Master's Thesis
Master of Theoretical and Computational Methods
Computational Aerosol Physics

Impact of Quantum Chemistry Parameters and Model Settings on Predicted Atmospheric Particle Formation

Vitus Besel

April 21, 2020

Supervisor(s): Jakub Kubečka, Theo Kurtén, and Hanna Vehkamäki

Examiner(s): Theo Kurtén and Hanna Vehkamäki

University of Helsinki
Faculty of Science

PL 64 (Gustaf Hällströmin katu 2a)
00014 Helsingin yliopisto

HELSINGIN YLIOPISTO — HELSINGFORS UNIVERSITET — UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Degree programme
Faculty of Science	Master of Theoretical and Computational Methods Computational Aerosol Physics

Tekijä — Författare — Author

Vitus Besel

Työn nimi — Arbetets titel — Title

Impact of Quantum Chemistry Parameters and Model Settings on Predicted Atmospheric Particle Formation

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidantal — Number of pages
Master's Thesis	April 21, 2020	47

Tiivistelmä — Referat — Abstract

We investigated the impact of various parameters on new particle formation rates predicted for the sulfuric acid - ammonia system using cluster distribution dynamics simulations, in our case ACDC (Atmospheric Cluster Dynamics Code). The predicted particle formation rates increase significantly if rotational symmetry number of monomers (sulfuric acid and ammonia molecules, and bisulfate and ammonium ions) are considered in the simulation. On the other hand, inclusion of the rotational symmetry number of the clusters only changes the results slightly, and only in conditions where charged clusters dominate the particle formation rate because most of the clusters stable enough to participate in new particle formation display no symmetry, therefore have a rotational symmetry number of one, and the few exceptions to this rule are positively charged. Further, we tested the influence of the application of a quasi-harmonic correction for low-frequency vibrational modes. Generally, this decreases predicted new particle formation rates, and significantly alters the shape of the formation rate curve plotted against the sulfuric acid concentration. We found that the impact of the maximum size of the clusters explicitly included in the simulations depends on the simulated conditions and the errors due to the limited set of clusters simulated generally increase with temperature, and decrease with vapor concentrations. The boundary conditions for clusters that are counted as formed particles (outgrowing clusters) have only a small influence on the results, provided that the definition is chemically reasonable and the set of simulated clusters is sufficiently large. We compared predicted particle formation rates with experimental data measured at the CLOUD (Cosmics Leaving OUtdoor Droplets) chamber. A cluster distribution dynamics model shows improved agreement with experiments when using our new input data and the proposed combination of symmetry and quasi-harmonic corrections., compared to an earlier study based on older quantum chemical data.

Avainsanat — Nyckelord — Keywords

Computational Aerosol Physics, Atmospheric Particle Formation

Säilytyspaikka — Förvaringsställe — Where deposited

Muita tietoja — Övriga uppgifter — Additional information

Acknowledgments

First of all, I want to thank the people who tirelessly corrected and reviewed this thesis and helped me shaping it into its final form: Jakub Kubečka, who endured my endless stream of questions patiently, guided me when I felt clueless and definitely puts the "fun" in "office", Theo Kurtén for his precious feedback and who I am especially grateful to for bringing me into this research group in the first place, and Hanna Vehkamäki for her personal as well as professional support and for creating such an enabling and healthy working environment.

Moreover, I want to thank all the other wonderful people of the Computational Atmospheric Physics group: Dina, Anna, Monica, Sarah, Valtteri, Roope, Golnaz, Bernhard, Tommaso, Olli, Stephen, Ivo and Matias. I am grateful for being able to work with you and the ways each of you contributes to a supportive research environment, but also for all the good time we spent outside of work making sure that life does not get boring.

Eventually, this research was supported by the European Research Council project 692891-Damocles, Academy of Finland, University of Helsinki, Faculty of Science ATMATH project. I also like to thank CSC - Finnish IT Centre for access to computer clusters.

Contents

List of Symbols

Variables (Latin alphabet):

symbol	*explanation*	*unit*
A	Surface area	m^2
B_{avg}	Average molecular moment of inertia	kg m^2
C	Concentration	cm^{-3}
	Correction term	cal/mol
E	Energy	a.u.
G	Gibbs free energy	kcal/mol
H	Enthalpy	kcal/mol
m	Mass	g
p	Pressure	Pa
P	Vapour pressure	Pa
r	Spatial coordinates	m
R_g	Radius of gyration	Å
S	Entropy	cal/mol
	Saturation ratio	
T	Temperature	K
V	Volume	m^3
ZPVE	Zero point vibrational energy	a.u.

Variables (Greek alphabet):

symbol	*explanation*	*unit*
β	Collision coefficient	m^3/s
γ	Evaporation rate	m^3/s
μ	Moment of inertia	kg m^2
ω	Frequency	1/cm
ρ	Electron density	
σ	Rotational symmetry number (symmetry context)	

	Temperature-dependent surface tension (classical liquid droplet)	N/m
τ	Spatial coordinates and electron spin	
ϑ	External potential	a.u.

Constants:

symbol	*explanation*	*value*
h	Planck's constant	$6.62607015{\cdot}10^{-34}$ J s
k_B	Boltzmann constant	$1.3806503{\cdot}10^{-23}$ J/K
R	Gas constant	8.3144626 J/(K mol)

Mathematical symbols:

symbol	*explanation*
$\widehat{\mathrm{H}}$	Hamilton operator
O	Order of scaling
Φ	Reference wavefunction
Ψ	Wavefunction
T	Cluster operator
$\widehat{\mathrm{T}}$	Kinetic energy operator
$\widehat{\mathrm{V}}$	Potential energy operator

List of Abbreviations

abbreviation	*full term*
ACDB	Atmospheric Cluster Database
ACDC	Atmospheric Cluster Dynamics Code
BDE	Birth-death equations
CC	Coupled-Cluster method
CERN	*Conseil européen pour la recherche nucléaire* (European Organization for Nuclear Research)
CLOUD	Cosmics Leaving OUtdoor Droplets
D	Double excitations
DFT	Density Functional Theory
DLPNO	Domain-based local pair natural orbital
GM	Global minimum
HF	Hartree-Fock method
HKT	Hohenberg-Kohn theorem
IPCC	Intergovernmental Panel on Climate Change
JKCS	Jammy Key for Configurational Sampling
LM	Local minimum
MS	Mass spectrometry
PES	Potential energy surface
RRHO	Rigid Rotor-Harmonic Oscillator
S	Single excitations
T	Triple excitations

Introduction

Rapid climate change is one of the most pressing issues of our time. It is a general consequence of humanity underestimating or straight out ignoring its impact on the environment. However, from the beginning of the 2000s climate change shifted more and more into the public focus and for large parts a rethinking occured, with the general public acknowledging that it poses a problem for all of humanity [Capstick et al., 2015].
For tackling climate change efficiently, it is necessary to understand all the climate related processes as well as possible. The climate is an immensely complex concept and it is impossible to accurately keep track of all issues included and correlated. The Intergovernmental Panel on Climate Change (IPCC) aims for organizing research and insights gained into climate change. It regularly publishes reviews and has already assessed the radiative forcing of different factors in 2013 [Myhre et al., 2013]. It is well known that greenhouse gases such as carbon dioxide and methane contribute to the global warming. On the other hand, the influence of aerosol-radiation and aerosol-cloud interactions remain largely uncertain, which is why they are of great interest for atmospheric researchers.

Air contains not only roughly $2.64{\cdot}10^{19}$ cm^{-3} molecules, but can also contain up to millions of aerosol particles in the same volume. Aerosols are a suspension of liquid, solid or amorphous particles in the air [Hinds, 1999, Virtanen et al., 2010] and range in size from nanometers to several hundreds of micrometers. By scattering and absorbing radiation from the sun they have a direct impact on the climate. Furthermore, aerosol particles are essential for cloud formation processes since they are required as seeds for cloud droplet formation, and therefore have significant

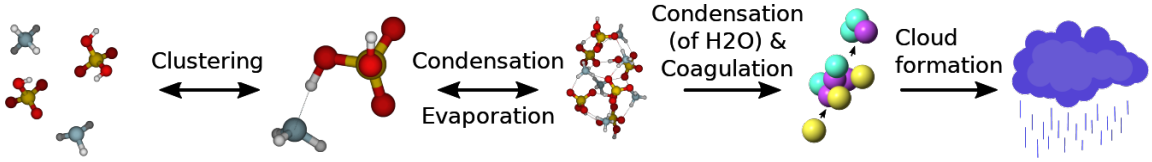


Figure 1: An illustration of the growth process of gaseous molecules up to cloud formation.

indirect impact onto the climate too. So-called *primary* aerosols are particles which are directly emited to the atmosphere from natural or anthropogeneous sources. *Secondary* aerosols form from gaseous molecules in the atmosphere, as illustrated in Fig. 1 [Brock et al., 1995, Kulmala et al., 2004, Clarke, 1992]. It has been estimated that 40 to 70 % of all cloud condensation nuclei are formed by this new particle formation [Yu and Luo, 2009, Wang and Penner, 2009, Merikanto et al., 2009], putting this process into the spotlight for atmospheric research [Mäkelä et al., 1997, Metzger et al., 2010, Schobesberger et al., 2013, Schobesberger et al., 2015]. As a sidenote it should be mentioned that new particle formation is commonly referred to as "nucleation": this is technically incorrect as the process may not always involve energy barriers.

A large number of studies have been devoted to unraveling the precise molecular-level mechanisms responsible for the first steps of atmospheric aerosol formation [Jacobson et al., 2000, Sun and Ariya, 2006, Lohmann and Feichter, 2005, Mäkelä et al., 1997, Metzger et al., 2010, Schobesberger et al., 2015, Schobesberger et al., 2013]. However, due to the large number of potentially participating compounds, the process is exceedingly complex, and both experimental and modeling tools still require development in order to reach a quantitative understanding of atmospheric new particle formation.
Sulfuric acid and ammonia (cf. Fig. 2) have been identified to be key players in the formation of aerosols [Schobesberger et al., 2015] and have been studied extensively in the past both experimentally and theoretically. The main reason for this is that the equilibrium vapour pressure of sulfuric acid in the presence of bases, of which ammonia is the most abundant in the atmosphere, is very close to zero. Furthermore, in contrast to for example most low-volatility organic compounds, sulfuric acid forms

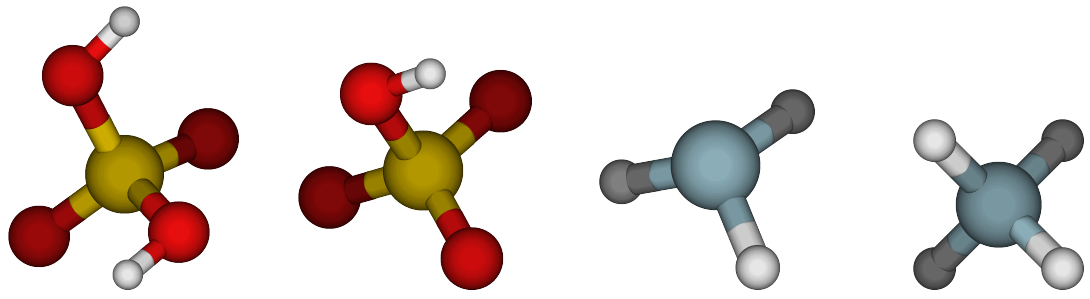


Figure 2: The present work investigates clustering and particle formation processes including sulfuric acid (a: H_2SO_4), its conjugate base bisulfate (b: HSO_4^-), ammonia (c: NH_3), and its conjugate acid ammonium (d: NH_4^+) with oxygen in red, sulfur in yellow, nitrogen in blue, and hydrogen in white.

practically in a single oxidation step from the very volatile precursor SO_2. The subsequent reaction steps involve high-concentration reactants, H_2O and O_2, and are thus essentially instantaneous. Mixing ratios of sulfuric acid in the lower atmosphere very seldom exceed ten parts per trillion (i.e. about $2x10^8$ molecules per cm^3) even in highly polluted conditions, and are often much lower. Ammonia mixing ratios, on the other hand, can reach tens of parts per billion (on the order of 10^{11} molecules per cm^3).

The process of single molecules forming new particles in the atmosphere is difficult to measure experimentally. Molecule and cluster concentrations are low, and therefore they have to be measured with mass spectrometry (MS). In order for molecules to be detectable by MS they have to be charged, but measurements have shown that the new particle formation in the atmospheric boundary layer is dominated by electrically neutral particles [Rose et al., 2015]. To complement experiments, computational studies are needed, and in order to capture all relevant factors, such as proton transfers, quantum chemistry tools are necessary: density functional theory (DFT) is able to give accurate information about cluster structures [Kurtén et al., 2007, Ortega et al., 2012, Leverentz et al., 2013], and recently high level of theory wavefunction methods have been employed to calculate binding energies of atmospheric clusters [Kildgaard et al., 2018, Myllys et al., 2016b]. The Atmospheric Cluster Dynamics Code (ACDC) [McGrath et al., 2012, Olenius et al., 2013b] is a model that simulates molecular cluster formation for a given set of compounds, and has been used in many atmospheric studies [Kürten et al., 2016, Zhang et al., 2018, Bork et al., 2014, Myllys et al., 2016a, Olenius et al., 2013a]. ACDC uses Gibbs free energies provided by quantum chemistry to calculate evaporation coefficients and collision rates from the kinetic gas theory to eventually compute formation rates of atmospheric clusters and trace the cluster formation pathways.

The purpose of this thesis is to revisit sulfuric acid – ammonia clustering in the light of recent advances in the modeling schemes [Kubečka et al., 2019, Myllys et al., 2016b]. We use a newly sampled, extended set of molecular cluster structures for the sulfuric acid – ammonia system and review the effect of some common approximations or assumptions on modeled particle formation rates. We also conduct a systematic study on the effect of different approximations used in computing thermodynamic properties and consideration of symmetry in quantum chemistry programs on the outcome of ACDC calculations. Moreover, we compare our modeled particle formation rates with an experiment from the CLOUD (Cosmics Leaving OUtdoor Droplets) chamber at CERN [Dunne et al., 2016] and previous modeling of this experiment [Kürten et al., 2016].

1. Theory

This theory chapter will introduce the theoretical framework and tools used in this work, including the basics of quantum chemistry ranging from wavefunction methods over density functional theory to corrections for various inaccuracies in conventional methods. Eventually, an atmospheric application of the introduced tools is described in the sections "Configurational Sampling" and "Evolution of Atmospheric Cluster Populations".

1.1 Quantum Chemistry

1.1.1 Schrödinger Equation

A system can be characterized exactly by solving the time-independent Schrödinger equation

$$\widehat{\mathrm{H}}|\Psi\rangle = E|\Psi\rangle. \tag{1.1}$$

This equation is an eigenvalue equation with the Hamilton operator $\widehat{\mathrm{H}}$, the wavefunction Ψ and the electronic ground state energy of the system E. $\widehat{\mathrm{H}}$ represents the total energy, which consists of the kinetic energy of electrons $\widehat{\mathrm{T}}_\mathrm{e}$, the kinetic energy of nuclei $\widehat{\mathrm{T}}_\mathrm{N}$ and the potential energy operators for the electron-electron interactions $\widehat{\mathrm{V}}_\mathrm{ee}$, for the electron-nucleus interactions $\widehat{\mathrm{V}}_\mathrm{eN}$ and for the nucleus-nucleus interactions $\widehat{\mathrm{V}}_\mathrm{NN}$

$$\widehat{\mathrm{H}} = \widehat{\mathrm{T}}_\mathrm{e} + \widehat{\mathrm{T}}_\mathrm{N} + \widehat{\mathrm{V}}_\mathrm{ee} + \widehat{\mathrm{V}}_\mathrm{eN} + \widehat{\mathrm{V}}_\mathrm{NN}. \tag{1.2}$$

The Hamiltonian acts on the wavefunction Ψ and gives, if Ψ is an eigenfunction of $\widehat{\mathrm{H}}$, the eigenvalue E as result. Obtaining the energy of the system E and its wavefunction Ψ provides the possibility to deduce chemical properties of the system.

1.1.2 Wavefunction Methods

In principle, the Schrödinger equation (Eq. 1.1) is exact. In practice, it can only be solved analytically for few simple systems and various approximations have to be used for more complex ones. The most fundamental approximation applied, the

Born-Oppenheimer approximation [Born and Oppenheimer, 1927], is to separate the Schrödinger equation into an electronic part and a separate equation for the nuclei. It assumes that the motion of nuclei relative to the faster electrons is negligible, and consequently $\widehat{\mathrm{T}}_{\mathrm{N}}$ and $\widehat{\mathrm{V}}_{\mathrm{NN}}$ are trivially calculated. All following methods are solely concentrating on obtaining the *electronic* energy. One way of calculating the electronic energy is using wavefunction-based approximations, also called *ab initio* methods, because they do not require any other input than general physical constants and the structure/composition of the investigated system.

The most basic of wavefunction-based methods is the Hartree-Fock (HF) method [Hartree, 1928, Fock, 1930], which calculates 99% of the energy of a system. However, in order to reach chemical accuracy that residual 1%, also called electron correlation, has to be approximated as accurately as possible. The HF energy mostly holds contributions from the chemically irrelevant core electrons, whereas, the electron correlation is needed for describing bonding coming from chemically relevant energy differences. In HF each electron moves in the average field created by the other electrons. Each electron is fixed inside its orbital, which is the mathematic function enabling the calculation of the probability to find an electron in a certain region. In reality, an electron can move into another, unoccupied orbital, e.g. when they get energetically excited. These unoccupied, also *virtual*, orbitals and the excitation into them can be considered in the calculation of the energy and yields the residual 1% of the energy. Using these principles the HF method has been extended to more sophisticated methods, such as the Møller-Plesset perturbation theory [Møller and Plesset, 1934], Configuration Interaction or the Coupled Cluster method used here.

Coupled Cluster Method

The Coupled-Cluster (CC) method is a mathematically elegant way of including electron excitations in the calculation. The Coupled-Cluster wavefunction is

$$|\Psi\rangle = e^{\mathrm{T}}|\Phi\rangle \tag{1.3}$$

where Φ is the reference wavefunction and T is the cluster operator

$$\mathrm{T} = \mathrm{T}_1 + \mathrm{T}_2 + \mathrm{T}_3 + \cdots + \mathrm{T}_n \tag{1.4}$$

with T_1 being the operator for single excitations (S), T_2 for all double excitations (D), T_3 all triple excitations (T) up to n times excitation [Jensen, 2007].
In practice, inclusion of all possible excitations is computationally too expensive. Therefore, T is truncated after an order of choice and e^{T} is calculated as

a Taylor series. E.g. a Coupled Cluster calculation including single and double excitations is called CCSD. A popular approach is to additionally include triple excitations solely in terms of their perturbative impact on the single and double excitations, denoted as CCSD(T). Due to its favourable trade-off of computational cost and accuracy it is called the *gold standard* of computational chemistry [Klopper et al., 1997, Constans et al., 2000, Pitoňák et al., 2006].

CCSD(T) scales with $O(N^7)$ where N is the number of electrons in the system studied. $O(N^7)$ scaling is still too much computational cost for many applications, however, with the *domain-based local pair natural orbital* (DLPNO) CCSD(T) [Riplinger and Neese, 2013, Riplinger et al., 2013] there is a linearly scaling variant of the method available.

Wavefunction methods are hierarchic, meaning they can be systematically improved by including more and more possible excitations. This makes it easy to compare different methods in terms of accuracy.

The present work uses the ORCA program [Neese, 2018] for all DLPNO-CCSD(T) calculations.

1.1.3 Density Functional Theory (DFT)

A popular alternative to wavefunction methods is Density Functional Theory (DFT). In DFT the system is described by using the electron density instead of a wavefunction. This has the advantage that a calculation only depends on the three spatial coordinates of the system instead of the three spatial coordinates plus the electron spin for each electron of the system. In practice, the electron density in commonly used DTF methods is expressed in terms of one-electron orbitals just like in wavefunction methods and therefore end up with three spatial coordinates and the spin for each electron. However, DFT methods are able to capture a part of the electron correlation while they scale similarly to HF with respect to the number of electrons, where the precision and scaling depends on the used DFT functional.

The electron density ρ is connected to the electronic wavefunction by

$$\rho = N \int |\Psi(\tau_1, \cdots, \tau_N)|^2 d\tau_1, \cdots, d\tau_N, \tag{1.5}$$

where τ is the combined spatial coordinates of an electron r and its spin and N is the number of electrons. The DFT energy, with Born-Oppenheimer approximation

[Born and Oppenheimer, 1927] applied, is

$$E_{\mathrm{DFT}} = E_{\mathrm{T}} + E_{\mathrm{eN}} + E_{\mathrm{J}} + E_{\mathrm{XC}}. \tag{1.6}$$

Here, the electron-electron interaction E_{J} and the electron–nucleus interaction E_{eN} are trivially calculated as Coulomb interactions. The kinetic energy of the electrons E_{T}, the exchange-correlation term E_{XC} are more troublesome to calculate, and for these two terms different functionals have been developed, which are commonly based on empirical data or parameters. The core of DFT are the Hohenberg-Kohn theorems (HKT) [Hohenberg and Kohn, 1964]:

1. HKT: The external potential $\vartheta(r)$, thus the total energy of a system, is an unique functional of the electron density $\rho(r)$.

2. HKT: The density that minimizes the total energy is the exact ground state density. Therefore, DFT energies can be calculated variationally.

DFT energy is commonly approximated with the Kohn-Sham method [Kohn and Sham, 1965], which uses the kinetic energy of the electrons calculated easily in the wavefunction-based Hartree-Fock method $E_{\mathrm{T}}^{\mathrm{HF}}$. The result is the Kohn-Sham energy E_{KS}

$$E_{\mathrm{KS}}[\rho] = E_{\mathrm{T}}^{\mathrm{HF}}[\rho] + E_{\mathrm{J}}[\rho] + E_{\mathrm{eN}}[\rho] + E_{\mathrm{XC}}[\rho], \tag{1.7}$$

where

$$E_{\mathrm{XC}}[\rho] = (E_{\mathrm{T}}^{\mathrm{exact}}[\rho] - E_{\mathrm{T}}^{\mathrm{HF}}[\rho]) + (E_{\mathrm{ee}}[\rho] - E_{\mathrm{J}}[\rho]). \tag{1.8}$$

The exchange-correlation energy defined by Eq. 1.8 holds the discrepancy between the (unknown) *real* kinetic energy $E_{\mathrm{T}}^{\mathrm{exact}}$ and the calculated kinetic energy $E_{\mathrm{T}}^{\mathrm{HF}}$ and the discrepancy between the (unknown) *real* electron-electron interactions E_{ee} and the calculated electron-electron interactions E_{J}. E_{XC} is approximated with an DFT functional.

There are functionals ranging from basic ones, where E_{XC} solely depends on ρ, such as the VWN functional [Vosko et al., 1980], to functionals making use of derivatives of ρ like PW91 [Perdew and Wang, 1992] to functionals utilizing the exchange term calculated in HF, e.g., B3LYP [Stephens et al., 1994]. Functionals making use of the HF exchange term are called hybrid functionals. DFT methods commonly have problems in describing long-range correlation, or dispersion, correctly and have to be corrected accordingly. The present work uses the ωB97X-D functional [Chai and Head-Gordon, 2008]. This is a hybrid functional which is long-range corrected to perform well on long-range interactions present in atmospheric molecular cluster [Peverati and Truhlar, 2014, Kildgaard et al., 2018].

All DFT calculations were conducted with the Gaussian16 program [Frisch et al., 2016].

1.1.4 Basis Sets and Semi-empirical Methods

The wavefunction Ψ is central to wavefunction methods, but also essential for Kohn-Sham DFT methods. Generally, Ψ is initially unknown for any system and is approximated by a basis set created by combining basis-functions in the form of Gaussian distributions. Basis sets range from such assigning a single basis funtion to each orbital to basis sets including diffuse functions for flexibility or polarization functions. The larger the basis set, the higher is its accuracy, if chosen appropriately, and the larger is the computational cost of the calculation. This work uses the *valence double-zeta* 6-31++G** basis set for DFT calculations. That means that valence electrons, which take part in bonding, receive a more precise description through more basis functions than chemically inert core electrons. In this case each core electron basis function is comprised of six primitive Gaussian functions, whereas each valence electron is described by two basis functions of which one is comprised of a linear combination of three primitive Gaussian functions and the other one only of a single primitive Gaussian function. This is a medium sized basis set for DFT methods. We use the quite extensive augmented correlation-consistent polarized triple-zeta (aug-cc-pVTZ) basis set [Kendall et al., 1992] in DLPNO-CCSD(T) calculations.

Semi-empirical methods are methods based on the HF formalism, but they do not allow for a choice of basis set and replace some of the HF terms with parametrizations based on experimental data such as dipole moments, or ionization energies. This lowers the computational cost, while it makes accuracy heavily dependent on the difference between the computed system and the parametrization data set. Consequently, the outputs of these methods do not correspond to a traditional energy output and are not meaningful by themselves. Only energy differences computed with the exact same methods can yield sensible results. We use the semi-empirical method GNF-xTB [Grimme et al., 2017, Bannwarth et al., 2019] for the configurational sampling.

1.1.5 Thermodynamics

Quantum chemistry enables us to calculate various properties of chemical systems. Thermochemical properties, such as vibrational modes, are calculated using the Rigid Rotor-Harmonic Oscillator (RRHO). This approach approximates the behaviour of molecular systems as the behaviour of equilibrated ideal gas particles. Temperature is

included in terms of harmonic vibrations of the system around the equilibrium geometries and rigid rotation. In this thesis we are interested in thermodynamic properties of atmospheric molecules and molecular clusters. The Gibbs energy, also called Gibbs free energy,

$$G = H - T \cdot S \tag{1.9}$$

is the thermodynamic potential used for describing any system with a given temperature and pressure. T is the temperature and S is the entropy, and the enthalpy H can be calculated as

$$H = E_{\text{el}} + \text{ZVPE} + E_{\text{therm.corr.}}(\text{T}) \tag{1.10}$$

with the electronic energy E_{el}, the zero point vibrational energy ZVPE and a temperature-dependent thermal correction $E_{\text{therm.corr.}}$. ZVPE arises from vibrational motion of molecular systems even at 0 K.

A measure for the stability of molecular clusters is the difference of the cluster Gibbs free energy compared to the Gibbs free energies of the single molecules forming the clusters, the monomers:

$$\Delta G = G_{\text{Cluster}} - \sum_{i=1}^{N_{\text{Monomers}}} G_{i,\text{Monomer}}. \tag{1.11}$$

Respectively, the cluster formation enthalpy and entropy are defined as

$$\begin{aligned} \Delta H &= H_{\text{Cluster}} - \sum_{i=1}^{N_{\text{Monomers}}} H_{i,\text{Monomer}}, \\ \Delta S &= S_{\text{Cluster}} - \sum_{i=1}^{N_{\text{Monomers}}} S_{i,\text{Monomer}}, \end{aligned} \tag{1.12}$$

and the Gibbs formation free energy is related to these as

$$\Delta G = \Delta H - T \cdot \Delta S. \tag{1.13}$$

DLPNO-CCSD(T) yields a high accuracy. Due to computationally cost, however, we can only calculate energies for a given rigid clusters at a DLPNO-CCSD(T) level of theory, so called single point energies. On the other hand, with DFT it is feasible to calculate all thermodynamic properties. We use an approach combining the precision of DLPNO-CCSD(T) calculations and the speed of a DFT frequency analysis as

$$H_{\text{DLPNO},i} = H_{\text{DFT},i} + E_{\text{DLPNO},i} - E_{\text{DFT},i} \tag{1.14}$$

that has been shown yields high accuracy for the atmospheric clusters [Myllys et al., 2016b, Kildgaard et al., 2018]. Combining DFT with high-level wavefunction methods is a broadly accepted approach in quantum chemistry [Bauschlicher and Partridge, 1995, Montgomery et al., 1999, Baboul et al., 1999].

1.1.6 Corrections

RRHO does not accurately describe real molecular systems, which are not harmonic oscillators. Instead various corrections can be used to partially correct for some errors introduced by the RRHO.

Anharmonicity Scaling

First of all, it is difficult to solve the Schrödinger equation assuming a non-harmonic potential in larger molecular systems due to the computational cost. Typically, a scaling factor is derived for smaller systems and then applied to larger systems instead. This is done by multiplication of the harmonic frequencies with the scaling factor [Pople et al., 1993]. An anharmonicity scaling factor of 0.9960 derived by Myllys et al. [Myllys et al., 2016b] is applied to the modes with wavenumbers above 100 cm^{-1}.

Quasi-harmonic Approximation

Another flaw in RRHO for weakly bound clusters is that vibrations with low-lying frequencies would be better described as internal rotations [Grimme, 2012, Myllys et al., 2016b]. The vibrational entropy is

$$S_{\mathrm{V}} = R\left[\frac{h\omega}{k_B(e^{h\omega/k_BT}-1)} - \ln(1-e^{-h\omega/k_BT})\right] \tag{1.15}$$

with the gas constant R, Planck's constant h, frequency ω, the Boltzmann constant k_B and temperature T. The natural logarithm in Eq. 1.15 approaches ∞ for $\omega \to 0$, which leads to inaccurate results for low-lying frequencies. In this work, these inaccuracies are corrected for by replacing them with corresponding rotational entropy

$$S_{\mathrm{R}} = R\left[\frac{1}{2} + \ln\left(\sqrt{\frac{8\pi^3\mu' k_B T}{h^2}}\right)\right] \tag{1.16}$$

as suggested by Grimme [Grimme, 2012] with

$$\mu' = \frac{\mu B_{\mathrm{avg}}}{\mu + B_{\mathrm{avg}}} \tag{1.17}$$

where μ is the moment of inertia for a free-rotor and B_{avg} is the average molecular moment of inertia set to 10^{-44} kg m^2 restricting the effective moment of inertia to reasonable values [Grimme, 2012].
Eventually, the vibrational entropy is interpolated between S_{V} and S_{R} with a cut-off frequency of 100 cm^{-1} as suggested in the literature [Sure et al., 2014, Li et al., 2015]. The present work uses the Python script GoodVibes.py written by Funes-Ardois and Paton [Funes-Ardoiz and Paton, 2016] for applying the quasi-harmonic corrections.

Symmetry

The symmetry of molecules and clusters affects their entropy and thus also free energy. $C_{\mathrm{sym},i}$ is the term used for including symmetry in the entropy

$$C_{\mathrm{sym},i} = R \cdot T \cdot \ln(\sigma_i). \tag{1.18}$$

σ_i is the rotational symmetry number of system i, T is temperature, and R is the gas constant. This term is generally included automatically if symmetry is detected by the quantum chemistry program. However, this detection is based on geometrical thresholds, that are very tight. If these thresholds are slightly softened, some molecules or molecular cluster may be identified as symmetric using these new thresholds even though the quantum chemistry program did not recognize this. The program Symmol [Pilati and Forni, 1998] was used to identify symmetry and the symmetry-corrected Gibbs energy is:

$$\Delta G_i = G_i + C_{\mathrm{sym},i} - \sum(G_{\mathrm{monomer}} + C_{\mathrm{sym,monomer}}). \tag{1.19}$$

1.2 Configurational Sampling

A molecular atmospheric cluster of a certain composition can have numerous different configurations. This means that molecular clusters with the same composition but different with respect to translation, rotation or even (de)protonation of the monomers can be created (cf. Fig. 1.1). These clusters are characterized by different physical and chemical properties of which a measure of cluster stability is the energy needed to break the cluster apart (binding energy): the lower (more negative) the binding energy, the more stable is a cluster. Clusters that are stable in the atmosphere have favourable bonding networks. These energetically low-lying clusters are called *local minima* (LM) and can be pictured as valleys on a *potential energy surface* spanning all degrees of freedom within the system (cf. Fig. 1.2). The energetically lowest lying minimum is the *global minimum* (GM).

Configurational sampling scheme describes the procedure of searching for LMs, and eventually the GM. Here, the configurational sampling process was conducted with the Jammy Key for Configurational Sampling (JKCS) [Kubečka et al., 2019], that is a collection of scripts providing an easy way through a configurational sampling procedure by coupling of third party programs. JKCS was not used as a black box, but as a tool, whose in- and outputs at all stages of the process were carefully monitored and occasionally adjusted.

In the present work, we define the global minimum as the molecular cluster with

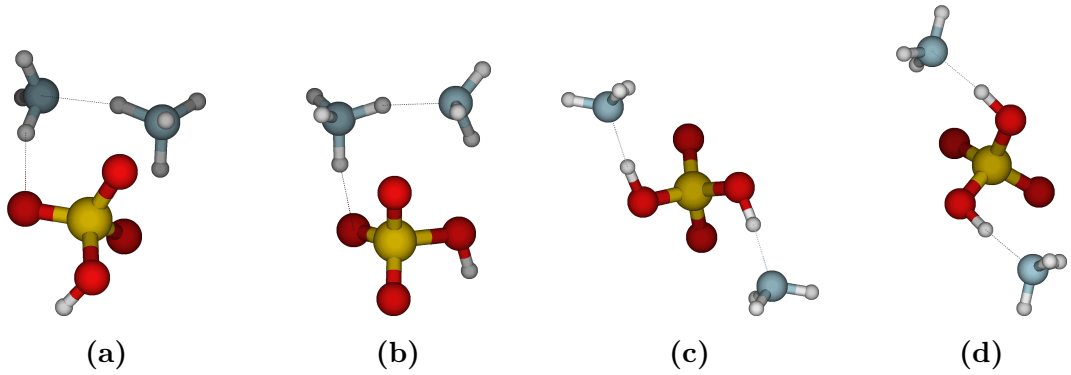


Figure 1.1: An example of diffferent configurations of an atmospheric cluster containing one sulfuric acid and two ammonium/ammonia molecules.

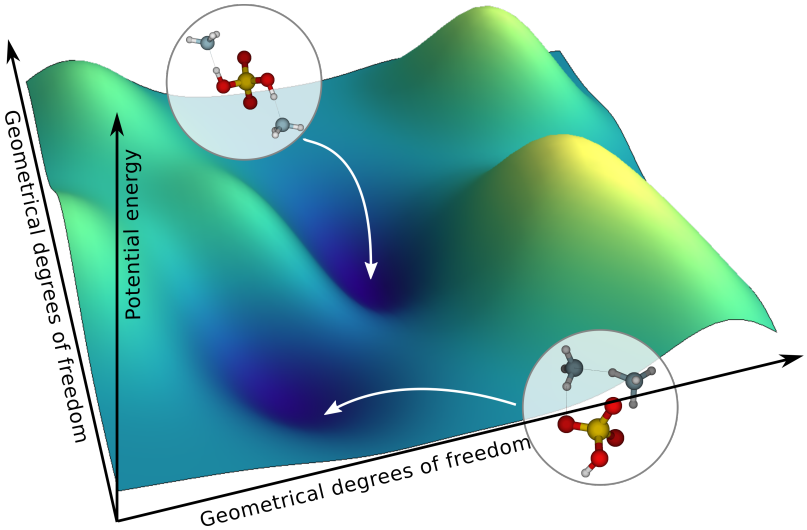


Figure 1.2: An illustration of a potential energy surface (PES) with two clusters depicted as example local minima.

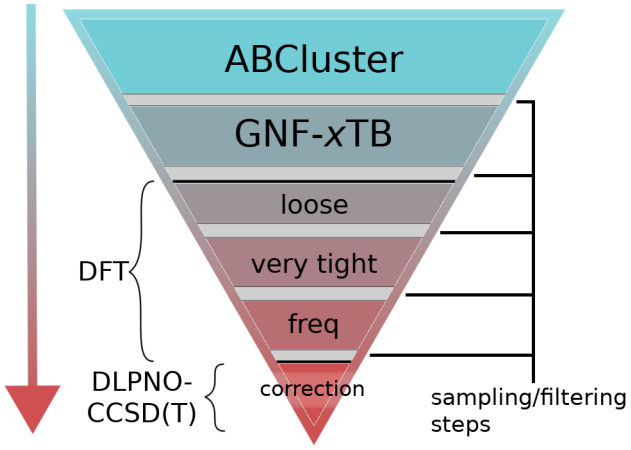


Figure 1.3: A scheme of the employed configurational sampling procedure. For DTF calculations optimization convergence criteria are first *loose*, then *very tight* and eventually vibrational modes are calculated (*freq*). All other used terms are explained in previous and following sections.

the lowest Gibbs energy (See 1.1.5). According to previous studies it is sufficient to focus on the global minimum instead of considering several local minima [Partanen et al., 2016, Kubečka et al., 2019]. However, most of our sampling process takes place in terms of the electronic energy, since it is cheaper to calculate. Nevertheless, one objective of an efficient configurational sampling procedure is to conduct these calculations on a minimal number of structures by using less precise, but also less expensive methods, and by taking after each level of theory only the structures which look most stable to the higher level of theory. Additionally a representative subset of all structures should be taken to the next higher level of theory, because the most stable structure at for example GNF-xTB level of theory is not necessarily the most stable structure at DFT level of theory. Even within DFT calculations with differently tight convergence criteria structures can have different relative stabilities, therefore it is important to take along this subset to not lose a potential global minimum.

Fig. 1.3 shows the sequence of methods employed in this work in order to find the global minimum efficiently: First of all ABCluster explores the configurational space and after rough filtering a subsequent calculation of GNF-xTB energies provides a first refinement of the structures. Afterwards, DFT calculations of increasingly tight convergence criteria followed by sampling and filtering rounds provide increasing accuracy and simultaneously reduce the number of structures studied. The final selection occurs after a frequency analysis at DFT level of theory which allows for a selection based on the DFT level Gibbs free energy. Eventually, a DLPNO-CCSD(T) provides a high accuracy electronic energy to be used in Eq. 1.14.

1.2.1 ABCluster

ABCluster is a program employing the artificial bee colony algorithm [Karaboga and Basturk, 2007] for molecular structure sampling [Zhang and Dolg, 2015, Zhang and Dolg, 2016]. The program is given the already optimized molecular structures of monomers that form the molecular clusters as input. ABCluster assumes that the monomers are rigid, therefore, all internal degrees of freedom of monomers are kept unchanged. This also prevents unwanted reactions during the exploration of the configurational space.
The artificial bee colony algorithm works by mimicking the behaviour of honey bee colonies. ABCluster implements the algorithm as the search for the best nectar as a food source, where the molecular cluster is the nectar and its energy is the quality of the nectar in terms of amount and distance from the hive. The adjustable parameters are numbers of different kinds of bees employed in the procedure: employed bees,

onlooker bees and scout bees. An iteration of the algorithm proceeds as

1. Random initial guesses for molecular clusters are created.
2. Employed bees look for good nectar sources: For each initial guess i another trial solution j is created by linear combination of a number of other guesses. The one of i and j with the higher potential energy is rejected.
3. Onlooker bees search for new nectar sources in proximity: Each onlooker bee selects a good solution k and a number of random other individual solutions. Then either a new trial solution is created by linear combination of k with the other individual solutions **or** by linear combination of the best one of the random selections with the random individual solutions. One of these possibilities is randomly chosen.
4. Scout bees look for new far nectar sources: Each current solution is examined and if it did not change for a number of cycles, a random new trial solution is created.
5. The algorithm continues at step 2 if a specified number of maximum cycles has not been reached yet.

By not always just accepting the lower energy the algorithm makes sure that it does not get stuck with few local minima, but actually explores the full potential energy surface.
The present work uses as input monomer structures optimized at a MP2/6-31++G(d,p) [Frisch et al., 1990a, Frisch et al., 1990b, Head-Gordon et al., 1988] level of theory to create 10 000 local minima for each monomer combination fulfilling the charge constraints. For the exploration of the PES, we assume that monomers are rigid, and thus, the cluster configuration is driven just by monomer-monomer Coulomb interactions and Lennard-Jones potential, which also accounts for atomic volume. To parametrize the monomers, we use the Mulliken partial charges from the MP2/6- 31++G(d,p) optimization and CHARMM (Chemistry at Harvard Molecular Mechanics) force field parameters of Lennard-Jones (6-12) equation [Vanommeslaeghe et al., 2010, Yu et al., 2012] for each atom. ABCluster used this parameterization to optimize structures after each step. We have performed 200 generations (loops) for each cluster combination incorporating a maximum of 5 scout bees.

1.2.2 Filtering and Sampling

Filtering refers to two procedures in the given context: firstly, the removal of cluster structures with significantly higher energies than the other structures that indicate

already highly unstable cluster structures. Secondly, the removal of structures which we assume to be identical within given thresholds. Sampling is in the given context the selection of a representative subset of all available structures with the general goal of reducing to number of calculations that have to be conducted. Filtering and sampling take place in terms of the energy computed at the respective levels of theory, the dipole moment and the radius of gyration R_g for characterizing the molecular clusters. The radius of gyration is calculated as

$$R_g^2 = \frac{\sum_{i=1}^{N} m_i |\overrightarrow{r_i} - \overrightarrow{r_{COM}}|^2}{\sum_{i=1}^{N} m_i}, \tag{1.20}$$

where m_i is the mass and $\overrightarrow{r_i}$ is the position of atom i for a total atoms of N in the cluster and $\overrightarrow{r_{COM}}$ is the center of mass of the cluster.
ABCluster creates thousands of molecular cluster structures of which many converge to the same structure once optimized by GFN-xTB and thus filtering has to be used. This reduces the amount of unecessary calculations drastically. For sampling we employ an algorithm selecting the lowest energy structure, then removing all structures within a certain radius based on the three variables. After, a new structure is selected the algorithm repeats [Chaudhuri, 1994]. This sampling is applied as shown in Fig. 1.3.

1.3 Evolution of Atmospheric Cluster Populations

1.3.1 Critical cluster

The ΔG given by Eq. 1.19 corresponds to the reference pressure at which the monomer cluster free energies are computed, typically 1 atm. The reference pressure cancels out in the calculation of evaporation rates, and thus does not actually affect cluster distributions. For illustrative purposes and by analogy to the classical liquid droplet model it is possible to calculate so-called "actual ΔG" values by correcting for the actual partial pressures of the monomers, which are typically much lower than 1 atm. The classical nucleation theory expression for the formation free energy in a one component system is

$$\Delta G_{i,\text{actual}} = -(i-1)k_B T \ln(S) + (A_i - A_1)\sigma, \tag{1.21}$$

where S is the saturation ratio, A_i is the surface area of a cluster consisting of i moecules and σ is the temperature-dependent surface tension. The saturation ratio S is P/P_{eq} with the vapor pressure of the nucleating compound P and the temperature-dependent saturation vapor pressure over a flat surface P_{eq}. The equation consists of: The first term is the difference of free energy coming from the formation of a liquid droplet from gas molecules. This term is negative if $S > 1$, therefore for supersatured vapor,

for which the gas-liquid phase-transition is energetically favorable. The second term describes the surface energy and is always positive. Fig. 1.4 shows the interplay of these terms and Fig. A.1 shows the actual ΔG surface at atmospheric conditions for the studied sulfuric acid – ammonia system. The maximum of the final curve marks the critical cluster size. For clusters growing past this size it is energetically favorable to keep growing, however, this does not mean that no evaporation occurs anymore.

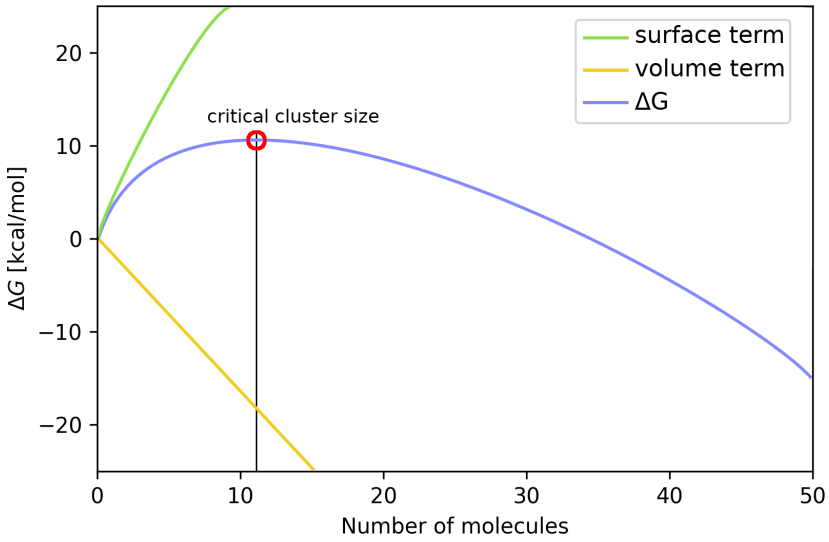


Figure 1.4: The cluster formation free energy and its surface and volume components according to the classical liquid droplet model.

1.3.2 Atmospheric Cluster Dynamics

There are two essential mechanisms of cluster dynamics: cluster form and grow by collision of molecules or clusters and subsequent attachment due to intermolecular interactions, such as van-der-Waals interactions, dipole-ion interactions or ion-ion interactions, or they fragment or evaporate single molecules. If we consider a cluster C forming from clusters/molecules A and B

$$A + B \rightarrow C \tag{1.22}$$

then the collision coefficient $\beta_{A+B\rightarrow C}$ for forming C is

$$\beta_{A+B\rightarrow C} = \left(\frac{3}{4\pi}\right)^{1/6} \left[6k_B T \left(\frac{1}{m_A} + \frac{1}{m_B}\right)\right]^{1/2} (V_A^{1/3} + V_B^{1/3})^2, \tag{1.23}$$

where m_X is the mass and V_X is the volume of cluster/molecule X. The volume is approximated using bulk liquid densities under the assumption of spherical clusters. On the other hand, cluster C can fragment/evaporate to A and B, which is described

by the evaporation rate

$$\gamma_{C\to A+B} = \beta_{A+B\to C}\frac{p_{\text{ref}}}{k_B T}\exp\left(\frac{\Delta G_{A+B\to C}}{k_B T}\right) \tag{1.24}$$

where $\Delta G_{A+B\to C}$ is the Gibbs free energy of formation of cluster C and the reference pressure p_{ref} is typically 1 atm. Furthermore, cluster C can be produced by external sources S_C, such as emissions, and can be removed by external losses L_C, such as coagulation onto pre-existing surfaces. All these processes are illustrated in Figure 1.5.

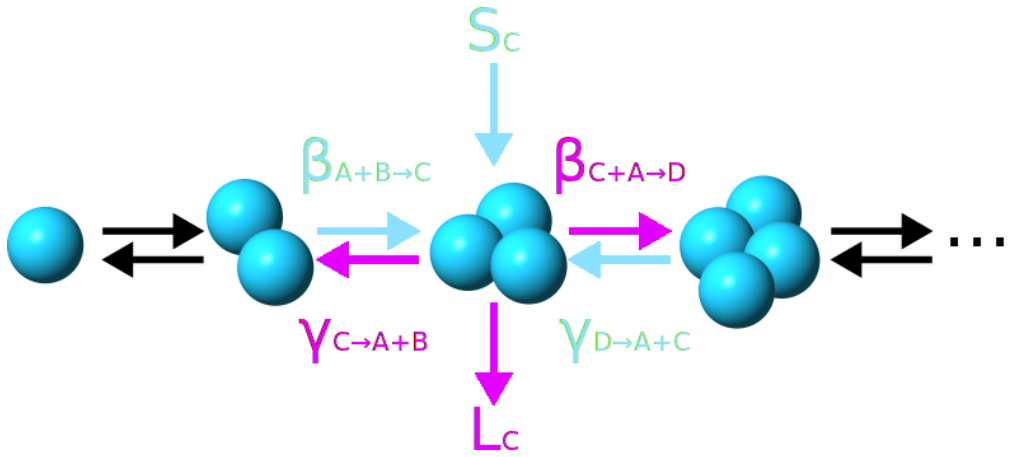


Figure 1.5: An illustration of cluster dynamics for an arbitrary homogeneous system. In light blue processes creating the central cluster C, and in purple processes removing cluster C, where S_C are external sources, L_C are external losses, γ are evaporation rates and β are collision rates.

Eventually all mentioned processes can be incorporated into the so-called birth-death equations (BDE) describing the behaviour of a population of clusters in time. The BDEs are solved by integrating the time-derivatives of cluster concentrations and for cluster C they take the form

$$\begin{aligned}\frac{d[C]}{dt} = &\sum_{C=A+B}^{A\leq B}(\beta_{A+B\to C}[A][B] - \gamma_{C\to A+B}[C]) + \\ &\sum_{C=D-A}^{A\leq C}(\gamma_{D\to A+C}[D] - \beta_{A+C\to D}[A][B]) + S_C - L_C[C],\end{aligned} \tag{1.25}$$

where $[X]$ is the concentration of X. All positive terms create C and all negative terms remove C.

Atmospheric Cluster Dynamics Code

The present work uses the Atmospheric Cluster Dynamics Code (ACDC) [McGrath et al., 2012, Olenius et al., 2013b], which generates the BDEs and solves

them explicitly by employing the `ode15s` [Shampine and Reichelt, 1997] solver in MATLAB. We run ACDC with constant monomer concentrations, which in effect assumes that the monomer concentrations are much higher than the concentration of any clusters. This is the case for H_2SO_4 and NH_3 under most atmospheric conditions as well as in the CLOUD experiment. These monomers are allowed to collide and form clusters. Once clusters have formed, they can collide and evaporate. When a cluster that is not explicitly included in the simulation is formed, it evaporates immediately unless it forms a predefined outgrowing cluster deemed stable as shown in Fig. 1.6. Any collision that forms a cluster in the outgrowing area is counted towards the particle formation rate. In a time independent steady-state simulation, the particle formation rate is the number of all outgrowing clusters forming per second per volume [$cm^{-3}s^{-1}$].

Outgrowing areas are determined by the direction in which we expect the clusters to keep growing if they leave the set of simulated clusters. This is estimated by the ratio of the collision rate to the evaporation rate: for example, if a cluster i has a high collision rate with sulfuric acid and a low overall evaporation rate, the i + sulfuric acid cluster (if it is outside of the set of simulated clusters), is a good choice for being in the outgrowing area. Also any cluster containing more sulfuric acid monomers and more ammonia monomers than cluster i is then in the outgrowing area. In practice, multiple of these areas can be defined.

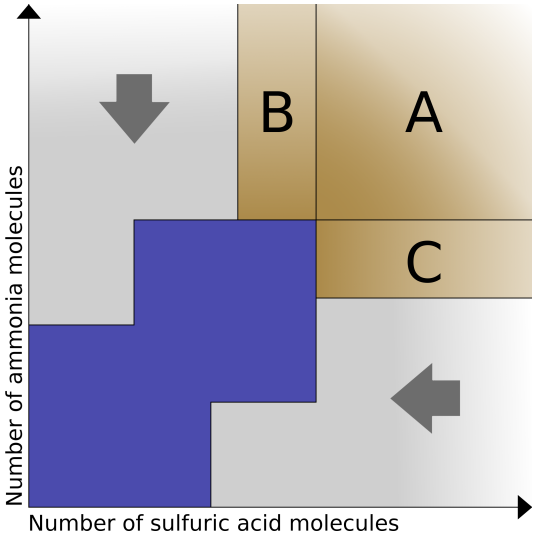


Figure 1.6: An illustration of cluster areas of the ACDC simulation: **Blue** is the set of simulated clusters, **brown** are possible areas of outgrowing clusters. If a cluster outside of these two areas (**grey**) is formed, it is brought back to the blue area by evaporating either single ammonia or sulfuric acid molecules. The outgrowing areas are defined by their lower-left corner and span an infinite region to the right and up.

The model includes parameters for explicit modeling of different types

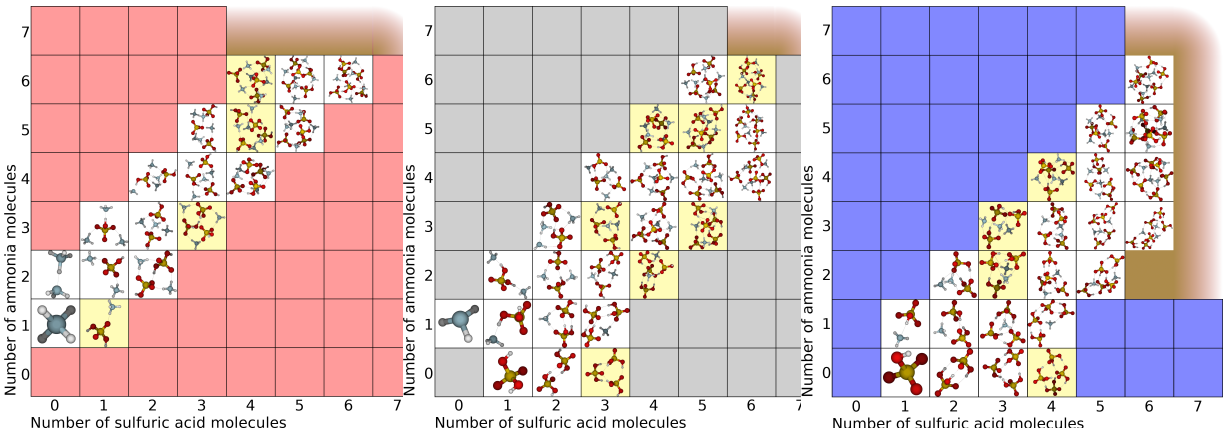
of external losses of the molecular clusters in the CLOUD chamber [Kupiainen-Määtä and Olenius, 2017]. We consider wall losses, coagulation losses, dilution losses and, additionally, we account for the enhancement of loss rates through electrostatic interactions between ions by using conventional values reported for the CLOUD chamber [Kupiainen-Määtä and Olenius, 2017]. Our simulated system consists of clusters up to $(H_2SO_4)_6(NH_3)_6$, $(H_2SO_4)_6(NH_4^+)(NH_3)_5$, and $(HSO_4^-)(H_2SO_4)_5(NH_3)_6$ (cf. Fig. 2.1). We consider only clusters that have been shown to be relevant for growth to larger sizes by Olenius et al. [Olenius et al., 2013a] and do not include hydration of clusters. The effect of hydration may have an impact on the overall performance of the model compared to the experimental results [Henschel et al., 2016], but this effect is outside the scope of this study. The model includes charging and recombination of clusters via generic charger ions O_2^- and H_3O^+, which transfer their charge upon collision.

We compared our ACDC simulations with experimental data from the CLOUD chamber [Dunne et al., 2016]. We took all data points with temperature $T \pm 1$ K for temperatures 292 K, 278 K, 248 K, 223 K, or 208 K. Since the experiment was aiming for an ionization rate of 3 $cm^{-3}s^{-1}$, we dropped all data points with an ionization rate outside of the interval 2–4 $cm^{-3}s^{-1}$. The sulfuric acid concentration in the simulations ranged from 10^5 to 10^9 $cm^{-3}s^{-1}$ and the ammonia concentrations were 2, 10, 100, or 1000 ppt (this corresponds to a range of $6.6{\cdot}10^7$ cm^{-3} to $2.5{\cdot}10^{10}$ cm^{-3} depending on temperature). These concentrations were chosen in accordance with Kürten et al. [Kürten et al., 2016] in order to stay within relevant concentrations for the CLOUD experiment and allow comparability. The temperature as well as the ionization rate in the simulation was set to the respective mean of the measured data points.

2. Results and Discussion

2.1 Simulation Set-up

Figure 2.1 shows the set of simulated clusters in ACDC. Each cluster type is represented by the global minimum structure, with respect to their Gibbs free energy at 298.15 K found in this study. Cluster types far from the diagonal in these diagrams have been omitted as they play a minor role in cluster fluxes [Olenius et al., 2013a]. Depending on whether we apply the quasi-harmonic correction, the structure representing the global free energy minimum may differ. Unless otherwise denoted, we use the global minima corresponding to the case with quasi-harmonic approximation applied. Generally, the global minimum structure may change with temperature. In the present study, this was the case only for five structures. We tested the effect of using different minima at each temperature, and found that the error incurred by using the 298.15 K minima at all studied temperatures is less than 0.09 kcal/mol. This difference is insignificant, and thus we used the 298.15 K minima at all temperatures.

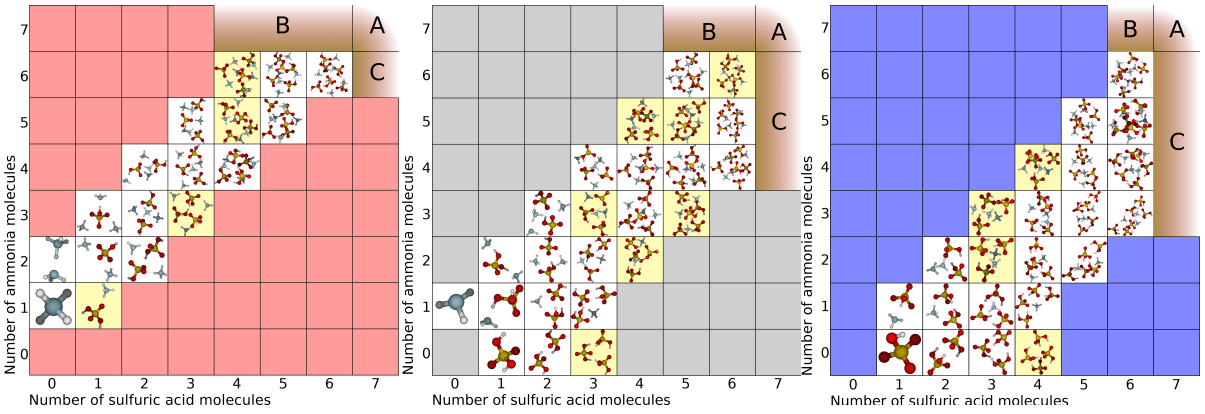


(a) Positive mode **(b)** Neutral mode **(c)** Negative mode

Figure 2.1: The global minima structures making up a set of simulated clusters. The brown area is the area in which formed cluster are counted towards the particle formation rate. The yellow highlighted clusters are clusters for which the global minimum structure changes depending on whether we apply the quasi-harmonic correction by the GoodVibes script or not.

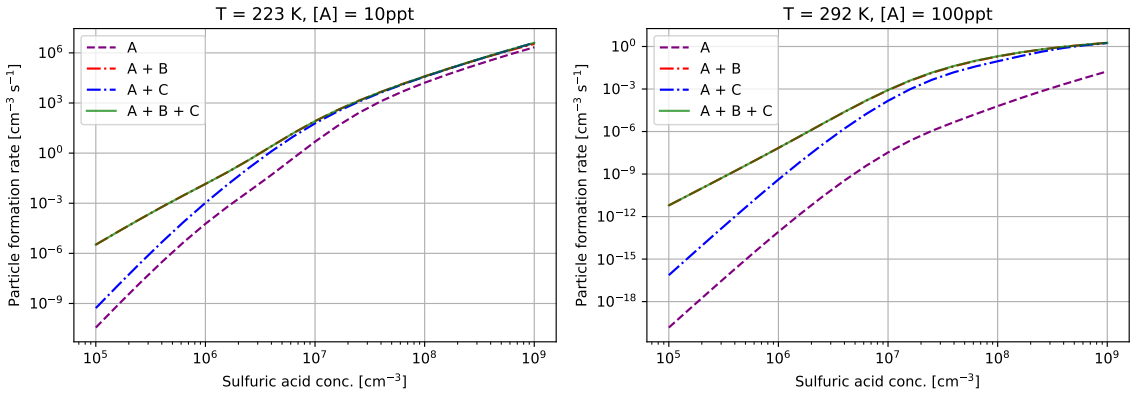
2.2 Determination of Boundary Conditions for Outgrowing Clusters

First, we want to estimate the impact of variations of boundary conditions for outgrowing clusters, i.e. the definition of outgrowing areas. We set up a series of simulations with successively extending outgrowing areas: area A, A+B, A+C, and A+B+C, as shown in Fig. 2.2. We tested each area for 292 K and an ammonia concentration of 100 ppt, and for 223 K and an ammonia concentration of 10 ppt with sulfuric acid concentrations ranging from 10^5 to 10^9 cm^{-3}. The resulting particle formation rates as a function of acid concentration are shown in Fig. 2.3. Outgrowth through area A produces the lowest rates because it cannot be reached by monomer addition, but only through collision with already formed clusters, e.g., $(H_2SO_4)(NH_3)$. The definition which only contains the outgrowing area A is thus a poor choice if monomer concentrations exceed cluster concentrations by several orders of magnitude. In all present cases, monomer concentrations are significantly higher than cluster concentrations and outgrowth occurs primarily through monomer addition. All other reasonable choices of boundary definitions show, especially in conditions of atmospherically relevance (roughly with particle formation larger than 0.01 cm^{-3}s^{-1}) negligibly small differences. The effect of including area B is larger than that of including area C as in most of the simulated conditions cluster encounter ammonia molecules much more often than sulfuric acid molecules.



(a) Positive mode **(b)** Neutral mode **(c)** Negative mode

Figure 2.2: The definition of the outgrowing areas (brown) for the test of the different outgrowth directions. The negative mode outgrowing area has been slightly reduced in comparison to Fig. 2.1 because including the row with two ammonia and six sulfuric acid molecules would automatically include area A, which is considered an independent area for this test.



(a) **(b)**

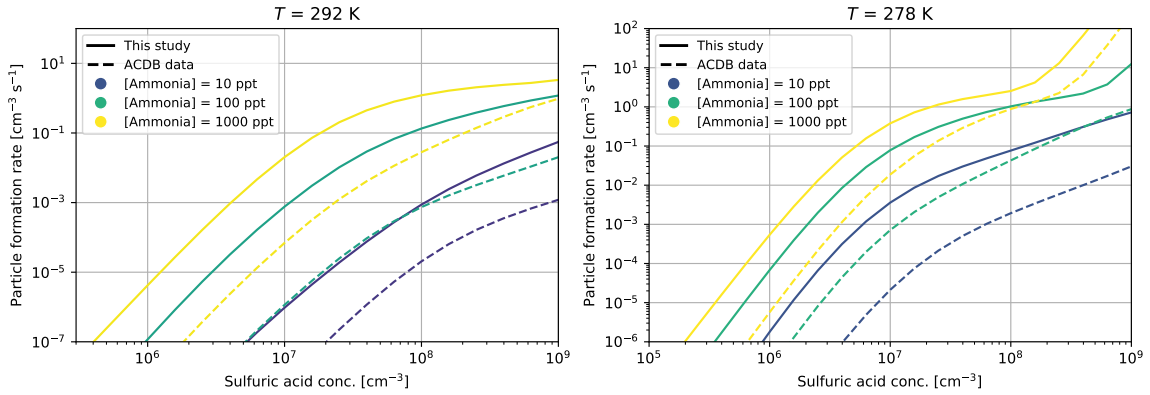
Figure 2.3: Particle formation rates from simulation runs with with different declarations of outgrowing areas. The (A+B+C) line is overlapping the (A+B) line.

2.3 Comparison to Previous Global Minima

We compared particle formation rates calculated using our new global minima to rates calculated using previously found global minima. We extracted the previous minima the from Atmospheric Cluster Database (ACDB) [Elm, 2019, Ortega et al., 2012, Loukonen et al., 2010] (November 2019). These data are based on quantum chemistry calculations, which generally do not consider symmetry, except for the ammonium molecule. We recalculated all the structures taken from ACDB with the DFT and DLPNO-CCSD(T) methods described in section 2.1. Not all molecular cluster types we sampled were available in the database, but also not all cluster types in the database have been sampled by us. Therefore, these comparison simulations were run using only these clusters which were available both in ACDB and in our resampling. We also adjusted the outgrowing areas to the used set of simulated clusters.

The two data sets only differed in the global minima of the molecular clusters, and consequently their ΔS and ΔH but were calculated at the same level of theory. The results shown in this section should not be compared to simulation runs presented in the following sections due to the differences in considered clusters and symmetry treatment.

A comparison of our structures with ACDB shows that we identified the same global minima only for the smallest clusters. Our structures have a generally lower ΔG, and especially for clusters with three or more sulfuric acid molecules, the differences are greater than 1 kcal/mol. Our structures have a higher Gibbs free energy only in five cases, of which the highest difference is 0.67 kcal/mol for the $(NH_3)_3NH_4^+(H_2SO_4)_2$ cluster. As Fig. 2.4 shows, these differences result in significantly higher particle formation



(a) **(b)**

Figure 2.4: Particle formation rates from simulations run with cluster structures from ACDB and this study.

rates when using the new data. Generally, this illustrates the importance of performing systematic and thorough conformational sampling. The new molecular cluster data will be updated to the ACDB.

2.4 Symmetry

We prepared three different sets of ΔGs as input for ACDC differing in their consideration of symmetry. In the first set, we made sure that the Gaussian program does not use any symmetry at all, in the second set, we only turned on symmetry for the monomers, and in the third set, we additionally turned on symmetry for clusters. Cluster symmetry was identified visually and with the SYMMOL [Pilati and Forni, 1998] program and we corrected for it according to Eqs. 1.18 and 1.19.
Regarding the Gibbs free energy of formation, the treatment of symmetry results in a difference of roughly 1 kcal/mol between *"no symmetry at all"* and *"only monomer symmetry"* for each addition of a $(NH_3)(H_2SO_4)$ pair into the cluster (cf. dashed lines in Fig. 2.5). Fig. 2.6 shows the molecular clusters for which the program Symmol found symmetry which was not found by the quantum chemistry programs due to tight classification thresholds. Cluster (b) is a good example for the difficulties of determining symmetry only by eye: it has a vertical symmetry axis.

Fig. 2.7 shows particle formation rates obtained from ACDC for different symmetry inclusion cases. Taking symmetry into account results in more stable clusters and thus higher formation rates. As Eq. 1.19 shows, accounting for the symmetry of monomers changes ΔG for all involved clusters leading to a significant increase in

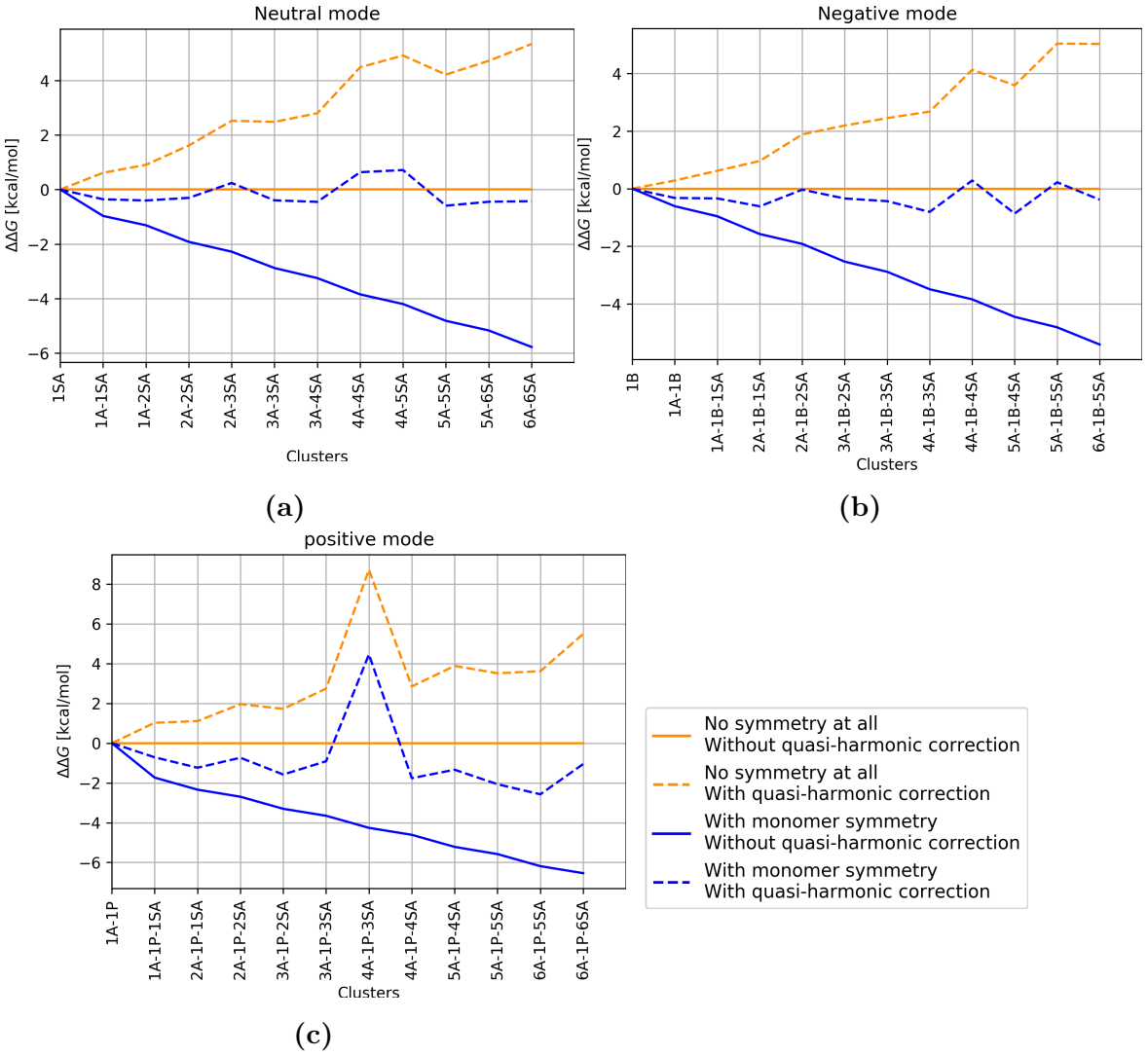


Figure 2.5: The differences of Gibbs free energies of formation $\Delta\Delta G$ in kcal/mol at 298 K and reference pressure 1 atm for (a) neutral clusters and (b) negative clusters with the same number of acid and ammonia molecules or one excess acid, and (c) positive clusters with the same number of acid and ammonia molecules or one excess ammonia: SA = sulfuric acid, A = ammonia, B = bisulfate, P = proton.

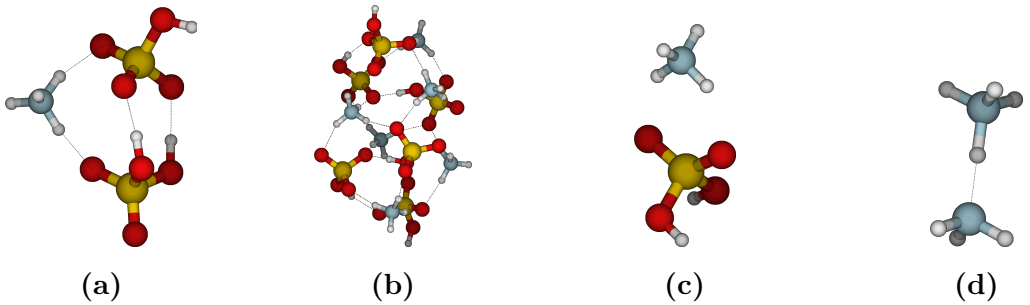


Figure 2.6: Molecular clusters with symmetry. Cluster (a) $(NH_4^+)(HSO_4^-)(H_2SO_4)$, (b) $(NH_4^+)_6(HSO_4^-)_6$ and (c) $(NH_4^+)(H_2SO_4)$ belong to the C_2 point group ($\sigma = 2$) and cluster (d) $(NH_4^+)(NH_3)$ belongs to the C_3 point group ($\sigma = 3$).

formation rates. On the other hand, the particle formation rates for *"only monomer symmetry"* and *"cluster and monomer symmetry"* differ only slightly at low concentrations of sulfuric acid and converge for higher concentrations. In order to explain this behaviour, we have to understand the shape of the particle formation rate vs. sulfuric acid concentration curve.

In general, the particle formation rate exhibits an inverted sigmoid shape. This shape is due to two different particle formation modes and a transition region between them. The first mode is in a region with a relatively high concentration of ions compared to that of sulfuric acid and ammonia, and the particle formation rate continuously increases with increasing sulfuric acid concentration. Then intermediate sulfuric acid concentrations correspond to a transition region where the particle formation rate increases only weakly with acid concentration. The second mode is found at even higher acid concentrations where the formation rate again increases strongly.

The cluster fluxes show that positive clusters grow into larger positive clusters, negative clusters grow into larger negative clusters, and that neutral clusters generally grow through collisions with charged monomers into larger negative and positive clusters at all monomer concentrations. Additionally, negative and positive clusters may recombine into neutral clusters. However, the growth of neutral clusters into larger neutral clusters is not observed unless the monomer concentrations are set high. Since the ionization rate stays constant, a higher concentration of sulfuric acid and/or ammonia means less charged clusters relative to neutral ones. Therefore, the fraction of purely neutral cluster growth increases with increasing monomer concentrations. The switch from charged clustering pathways to the neutral pathways gives rise to the two modes.

Fig. 2.6 shows that two small positively charged molecular clusters are symmetric, $(NH_4^+)(H_2SO_4)$ (c) and $(NH_4^+)(NH_3)$ (d), of which cluster (d) has the highest symmetry number σ of all clusters. Symmetry increases ΔG for these specific clusters located at the very beginning of the positive growth path. Therefore, clustering in the positive mode is influenced by the symmetry stronger than clustering in the neutral mode, and the differences caused by introducing cluster symmetry decrease with increasing sulfuric acid concentration as the importance of the neutral particle formation pathway increases. The positive mode particle formation rate increases when symmetry is taken into account as positive clusters evaporate back to cluster (c) and (d) at a lower rate, i.e., grow at a higher net rate. The differences in the particle formation rates caused by taking into account cluster symmetry are only significant for rather low formation rates of charged clusters and thus have little practical relevance in the boundary layer conditions.

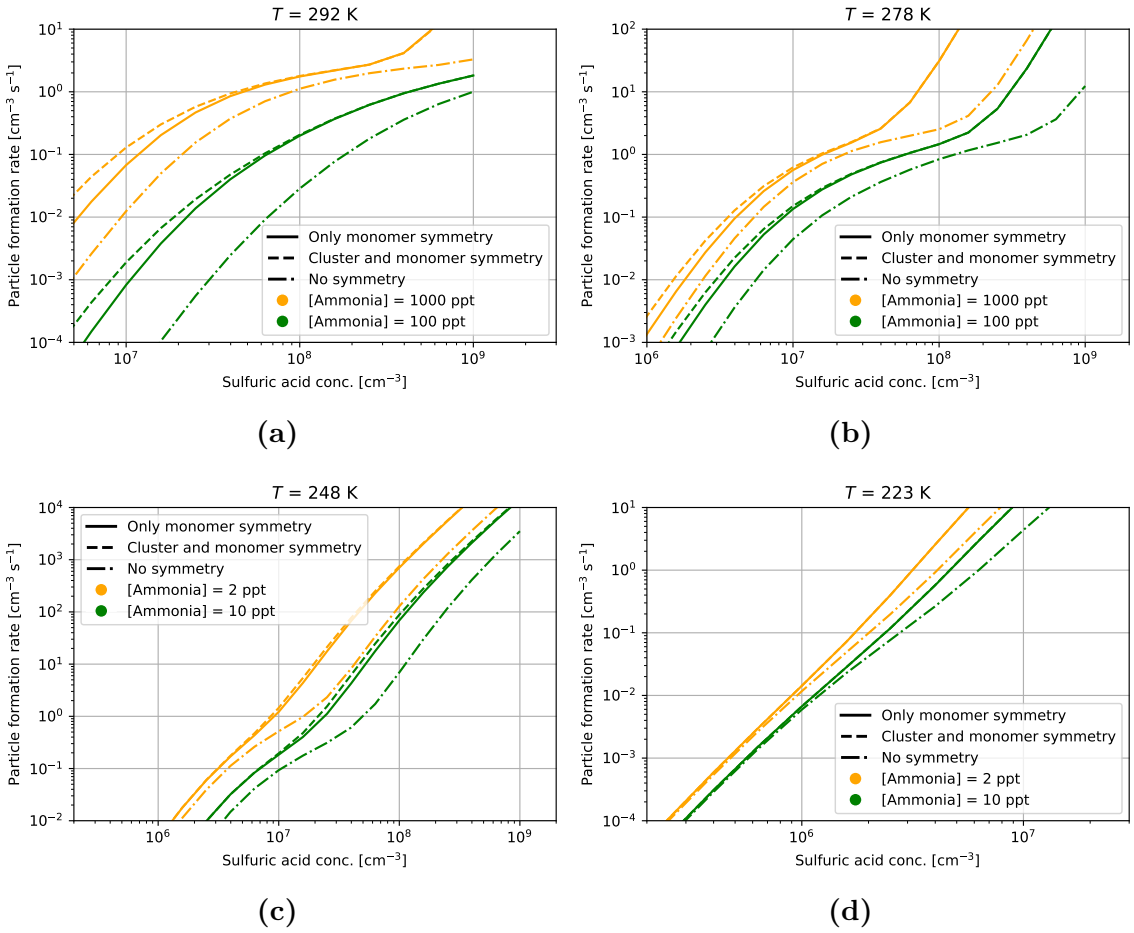


Figure 2.7: Particle formation rates in $\mathrm{cm}^{-3}\mathrm{s}^{-1}$ for different concentrations of sulfuric acid and ammonia at four different temperatures. Comparison of different choices of including symmetry.

2.5 Quasi-harmonic Correction

We prepared two different sets of ΔGs, one with the quasi-harmonic correction applied and one without it. The orange lines in Fig. 2.5 show clearly the difference of ΔGs for selected clusters in these two sets. While the quasi-harmonic correction increases all ΔGs, it affects some clusters more than others. The "4A-1P-3SA" cluster is an outlier and very strongly impacted by the quasi-harmonic correction due to the presence of an exceptionally low frequency mode at 14.24 cm^{-1}. The dashed blue line in Fig. 2.5 shows that the quasi-harmonic correction and the inclusion of monomer symmetry almost cancel each other out in terms of their effect on the ΔG where only in positive mode the decrease introduced by the symmetry is consistently larger than the increase introduced by the quasi-harmonic correction. It should be noted that although this cancellation is present for our sulfuric acid – ammonia system, the balance of the

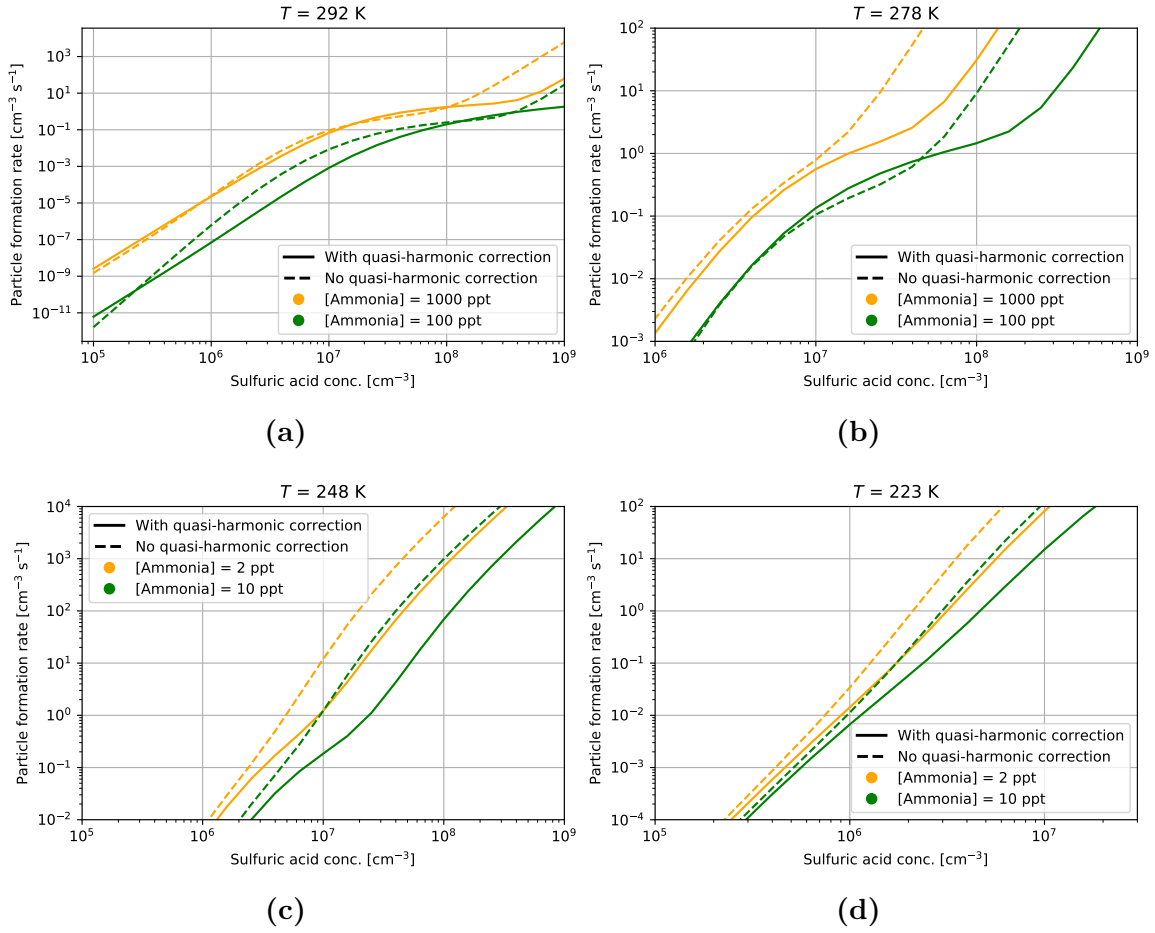


Figure 2.8: Particle formation rates in $cm^{-3}s^{-1}$ for different concentrations of sulfuric acid and ammonia at four different temperatures. The particle formation rates have been modeled with and without the quasi-harmonic correction. The ΔGs were calculated with monomer symmetry.

quasi-harmonic correction and the symmetry correction can be very different for other systems. Moreover, these two corrections are computationally very cheap, and thus do not constitute bottlenecks, and therefore we suggest that future studies perform both corrections rather than rely on uncertain error cancellation.
Charged clusters are, in general, more strongly bound than the corresponding neutral ones, as long as we compare clusters with roughly similar numbers of acid and base molecules. Therefore, neutral clusters tend to have more low vibrational frequencies. Subsequently, as shown in Fig. 2.8, the impact of the quasi-harmonic correction is stronger for higher sulfuric acid concentrations, where neutral clusters dominate the overall particle formation path. Overall, the quasi-harmonic correction decreases cluster stability. Thus, the quasi-harmonic correction slightly lowers the particle formation rate by shifting the curve to the right and down. This effect is more pronounced for higher temperatures, where the effect of entropy (including vibrational entropy) on free energy is greater.

2.6 Choice of the Set of Simulated Clusters

In this section, we investigate the influence of the set of simulated clusters on the particle formation rates. The 6x6 set of clusters is that shown in Fig. 2.1. The 5x5 set of clusters consists only of the clusters with less than six ammonia molecules or less than six sulfuric acid molecules. The 4x4 set of clusters consists only of the clusters with less than five ammonia molecules or less than five sulfuric acid molecules. The outgrowing areas are determined for each simulation set size using the same philosophy as in Fig. 2.1.

The results for different temperatures and ammonia concentrations are shown in Fig. 2.9. All the simulated particle formation rates converge with increasing sulfuric acid concentration analogous to the comparison of "only monomer" and "cluster and monomer" symmetries shown in Fig. 2.7. At the lower sulfuric acid concentration, where the particle formation rates differ for different simulation sets, clustering predominantly occurs through charged clusters. Generally, an increase of the cluster set size can either decrease the particle formation rate or leave it unchanged. Once a cluster grows into the outgrowing area, it is counted towards the particle formation rate. When the set size is increased, some outgrowing clusters of the previous, smaller set become "regular" clusters in the enlarged set. These – now "regular" – clusters can now either collide with other clusters/monomers or evaporate back to smaller clusters. The particle formation rate decreases if they evaporate at a significant rate.
Deeper insight is provided by the "actual ΔG"s of clusters which take the real partial

pressures of the monomers into account (cf. section 1.3.1). For most of the systems and conditions studied here, the actual ΔG surface contains one or more barriers between the monomers and the outgrowing clusters. The highest barrier on the lowest-energy path connecting the monomers to the outgrowing clusters (a saddle point on the actual ΔG surface) represents the so-called "critical cluster" as shown for a one component case in Fig. 1.4. Section 1.3.1 shows the actual ΔG according to the classical nucleation theory, however, the actual ΔG can also be obtained based on quantum chemistry calculations. All actual ΔG surfaces used in this work are based on quantum chemistry calculations. Due to the strong bonding between acids and bases, this lowest-energy path is typically close to the diagonal, i.e. the clusters on it tend to have roughly equally many acids and bases. Clusters smaller than the critical cluster are generally more likely to evaporate than to grow, while cluster larger than it are more likely to grow than to evaporate. (A common but false misconception is that clusters larger than the critical cluster do not evaporate at all.) For this reason, the explicitly simulated set of clusters should always include the critical cluster, and preferably some slightly larger clusters as well.
For charged clusters, the actual ΔG surface almost inevitably contains also a deep well (local minimum) corresponding to the core ion and a small number of neutral molecules ("ligands" in the language of co-ordination chemistry), located prior to the maximum corresponding to the critical cluster. The actual ΔG surface at 278 K, [ammonia] = 1000 ppt and [sulfuric acid] = 10^5 cm^{-3} is given as an example in the appendix A.1.

When extending the set of simulated clusters, there is a drop in the particle formation rates at all temperatures and ammonia concentrations for low sulfuric acid concentrations where cluster growth happens through charged modes, predominantly through the positive mode. The actual ΔGs in positive mode for the charged clusters with **four** sulfuric acid or **four** ammonia monomers are mostly lower or only slightly higher than those for the clusters with **five** sulfuric acid or **five** ammonia. Therefore, there is a clear drop in the formation rate since the set is extended to clusters that have significant evaporation rates. Going from the 5x5 set to the 6x6 set the added clusters have lower actual ΔGs than the largest clusters in the 5x5 set. Therefore, the formation rate drops only very little as clusters growing out of the 5x5 set also grow out of the 6x6 set. One exception is at 292 K and [ammonia] = 100 ppt for which the actual ΔG surface continuously rises in both the charged and the neutral modes (cf. Fig. A.2): when the cluster set is too small and does not contain the critical cluster the simulation can only give an upper bound for the particle formation rate, which will drop with each extension of the cluster set.

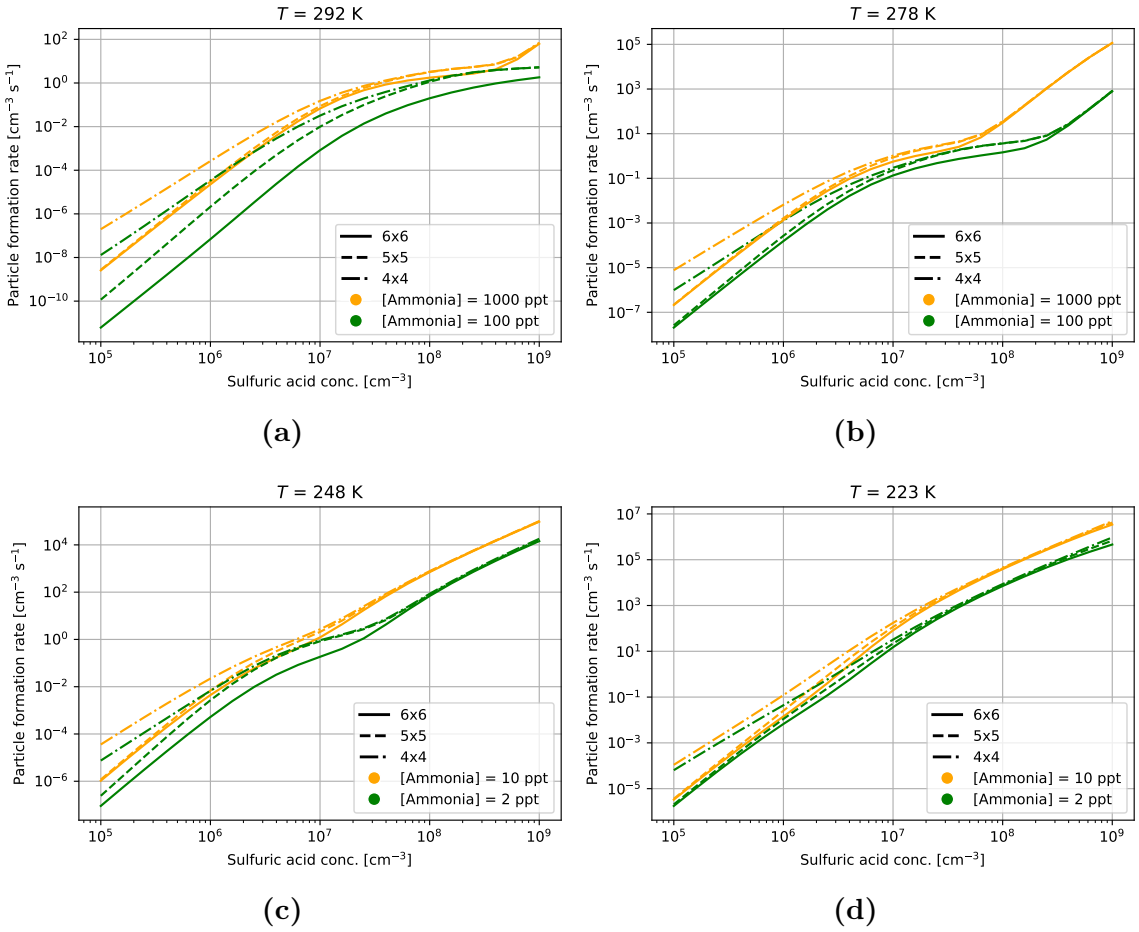


Figure 2.9: Particle formation rates in $\mathrm{cm}^{-3}\mathrm{s}^{-1}$ for different concentrations of sulfuric acid and ammonia at four different temperatures for various maximum sizes of clusters included in the simulation.

The particle formation rates in the transition region also decrease slightly when extending the set of simulated clusters. The reason for this is that the effect of the simulation set size is generally greater when there are multiple competing growth pathways, and thus multiple potential outgrowing clusters affected by the boundary conditions. Eventually, for high sulfuric acid concentrations cluster growth occurs through neutral clusters and in these cases the critical cluster is within the simulated cluster set at all temperatures and ammonia concentrations. Accordingly, the particle formation rates do not change when changing the set size. We conclude that in atmospherically relevant conditions a 5x5 cluster set is adequate for predicting the particle formation in the sulfuric acid – ammonia system.

2.7 Comparison to CLOUD

We plot our simulation results alongside the experimental data points selected and processed as described in the *Theory* chapter and alongside ACDC simulations reported by Kürten et al. [Kürten et al., 2016] based on quantum chemistry data from Ortega et al. [Ortega et al., 2012, Ortega et al., 2014]. The data from Ortega et al. corresponds to B3LYP/CBSB7 [Becke, 1993, Stephens et al., 1994] structures and vibrational frequencies combined with RICC2/aug-cc-pvtz energy corrections [Hättig and Weigend, 2000]. The outgrowing areas were chosen as in previous sections.

Some of the experimental data exhibit an inverted sigmoid shape with a plateau-like transition region consistent with the model runs. The present simulations match the experiment reasonably well for low temperatures and for moderate to high ammonia concentrations. This suggests that the presented model is especially suitable for cases

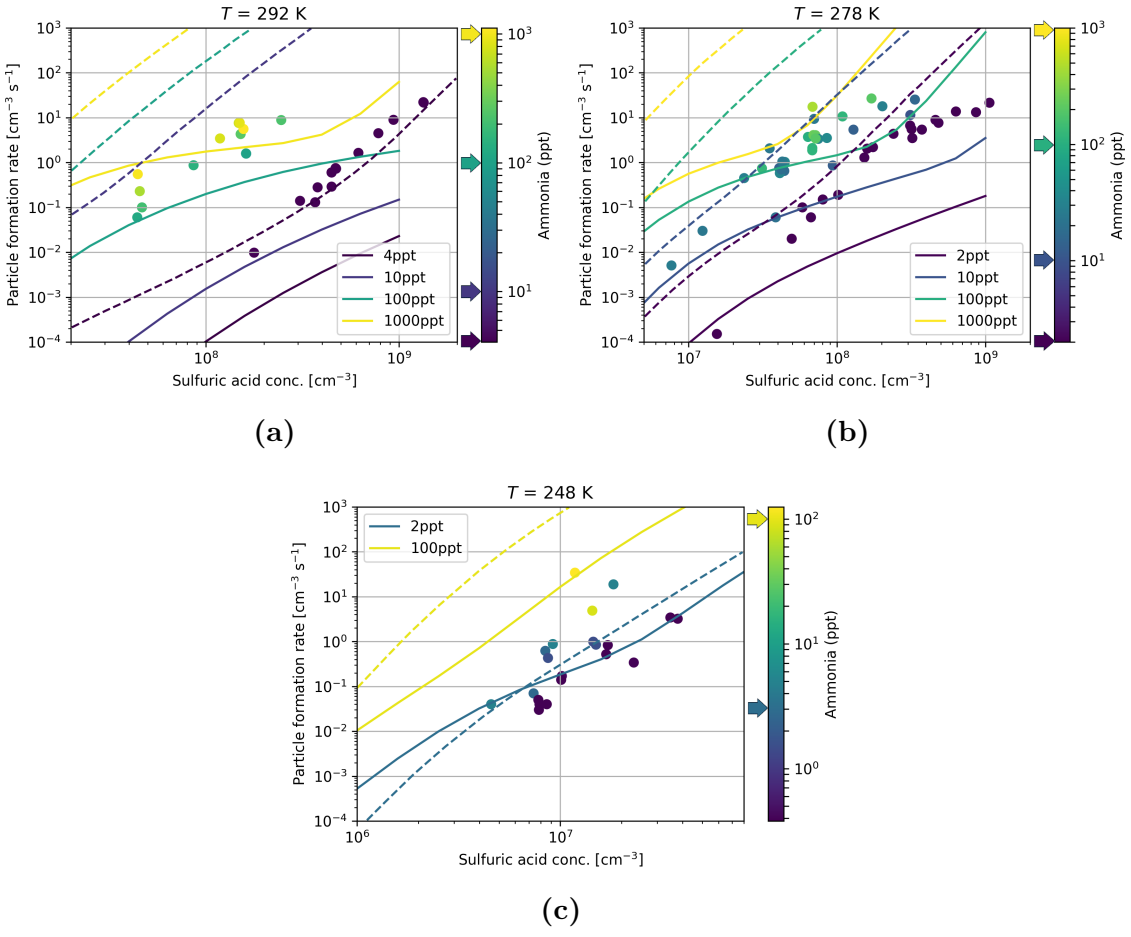


Figure 2.10: Particle formation rates in $cm^{-3}s^{-1}$ for different concentrations of sulfuric acid and ammonia modeled by us (full lines), Kürten et al. [Kürten et al., 2016] (dashed lines) alongside CLOUD measurements (dots) [Dunne et al., 2016]. Arrows on the color bar depict the ammonia concentrations where these simulations were conducted.

with low evaporation rates. For these cases, our model matches the experiments better than the ACDC simulations reported by Kürten et al., and generally displays particle formation rates up to several orders of magnitude lower than those in Kürten et al.'s. In contrast, the particle formation rates of Kürten et al. perform well for the lowest ammonia concentrations for which our new model is approximately one order of magnitude lower than that of Kürten et al.. The results of Kürten et al. are systematically higher than ours because their electronic energies in their ΔGs were calculated with RICC2 which is known to introduce overbinding.

3. Conclusion and Prospect

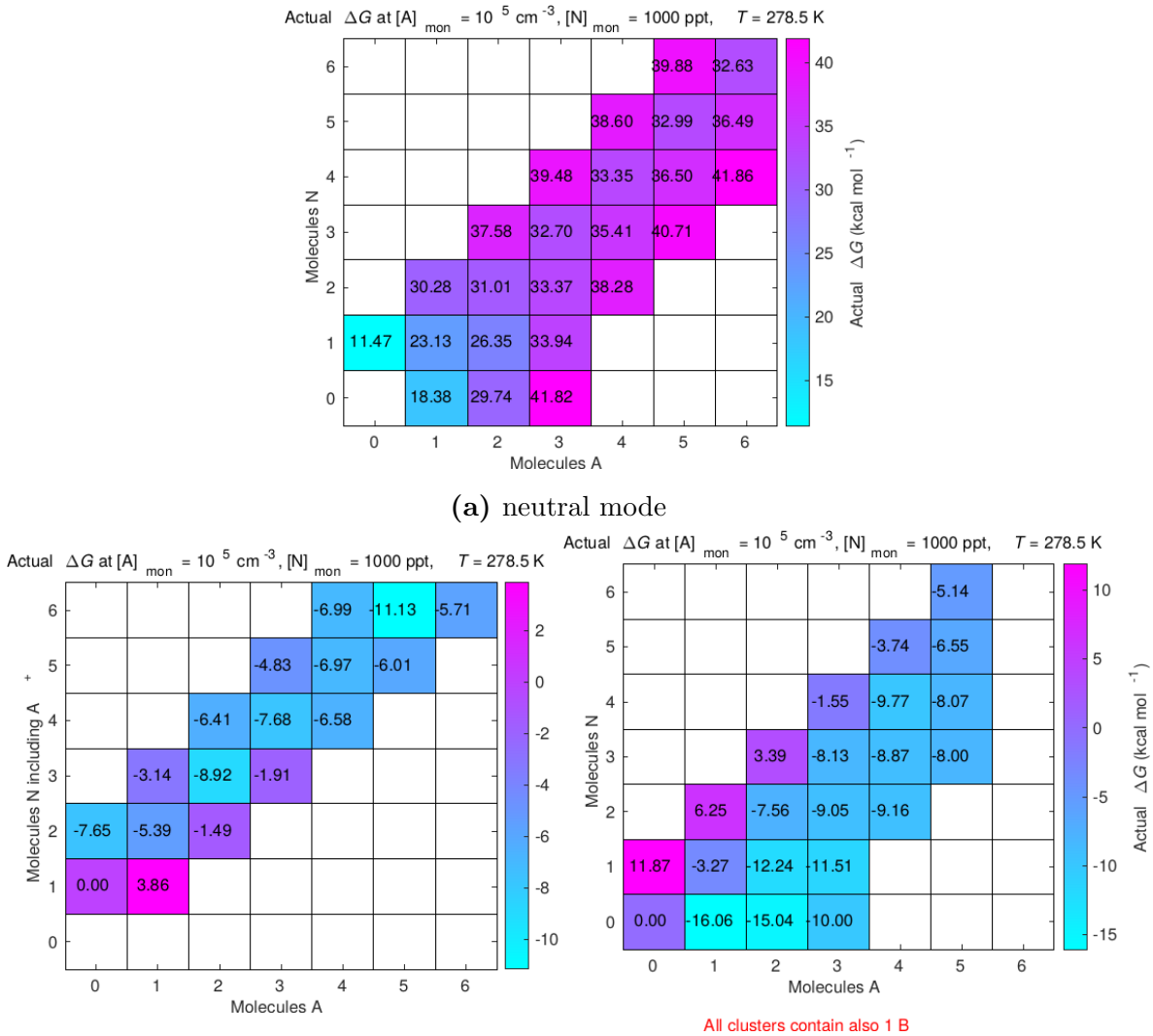
Our extensive systematic resampling for electrically neutral, positive and negative sulfuric acid - ammonia clusters with up to six acid and six base molecules resulted in significantly more stable cluster conformers than reported in previous studies for most cluster compositions. We used this set of clusters to explore the effect of various assumptions, approximations and boundary conditions on the particle formation rates predicted by the cluster distribution dynamics code ACDC.

First, we found that outgrowing clusters (i.e. which clusters are considered stable and thus counted toward the particle formation rate) — chosen in a chemically reasonable way and given a large enough set of simulated clusters is available — have a neglegible impact on predicted particle formation rates. Second, we showed that the consistent inclusion of the correct rotational symmetry number of all monomers increases particle formation rates significantly. Including rotational symmetry of the clusters, however, lead to little change and does not have to be accounted for in the sulfuric acid - ammonia system due to the relatively low symmetry of almost all minimum-energy clusters. We note that cluster symmetry may still be important for other systems, where clusters exhibit higher degrees of symmetry. Third, we investigated the effect of the quasi-harmonic correction for low-frequency vibrational modes. This correction leads to a general increase in ΔG attended by a change in the shape of the particle formation curve as a function of the sulfuric acid concentration, as it affects neutral clusters and thus neutral formation pathways more strongly than charged clusters and pathways. The quasi-harmonic correction and the inclusion of monomer symmetry give rise to error cancellation in their impact on the ΔGs to some degree, implying that previous studies which have typically omitted both corrections may nevertheless give adequate results. However, since neither correction is computationally expensive to perform, we strongly recommend that future studies carry out both. Fourth, we conducted simulations with cluster sets of differing size (6x6, 5x5 and 4x4). The results stress the importance of including the critical cluster (defined here as the highest barrier on the lowest-energy path connecting the monomers to the outgrowing clusters) within the explicitly simulated set of clusters. Accordingly, the minimum cluster set size required for a meaningful simulation is temperature and monomer

concentration depend as is the critical cluster composition. Furthermore, smaller cluster sets are sufficient to model conditions in which cluster growth is dominated by a small number of growth pathways, while larger sets are required to capture situations where multiple growth pathways contribute significantly.
We compared the particle formation rates predicted using our new cluster data set, and the optimal settings revealed by this work, to experimental results from the CLOUD chamber. Compared to previous ACDC simulations, we find substantially better agreement with experiments especially for lower temperatures and/or moderate to high ammonia concentrations.

Even though the results for the modeled experiment have improved, there is considerable room for even better agreement. This thesis gives a base for further refinement of the used methods, which will eventually expand atmospheric cluster dynamics simulations to models including more dimensional systems than just a binary one. The applied methods — especially configurational sampling — are very suitable for the incorporation of Machine Learning methods. Machine Learning has experienced a resurgence in the last decade and particularly data clustering methods are promising for increasing the efficiency of global minimum search for atmospheric clusters. With these and other advances in the fields of computer technology, atmospheric chemistry and data science predictions regarding particle formation and other aerosol phenomena will become more and more accurate. Eventually we will be able to understand the role of aerosols and cloud for the climate and will hopefully make educated decisions regarding the role humanity should play within nature.

Appendix A. Actual ΔG surfaces

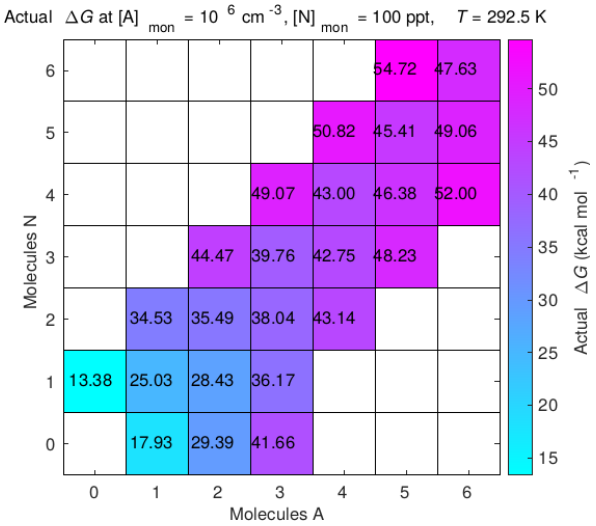


(a) neutral mode

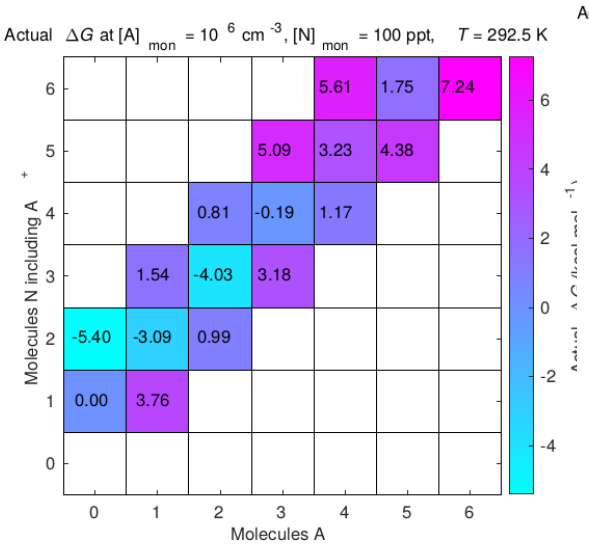
(b) positive mode

(c) negative mode

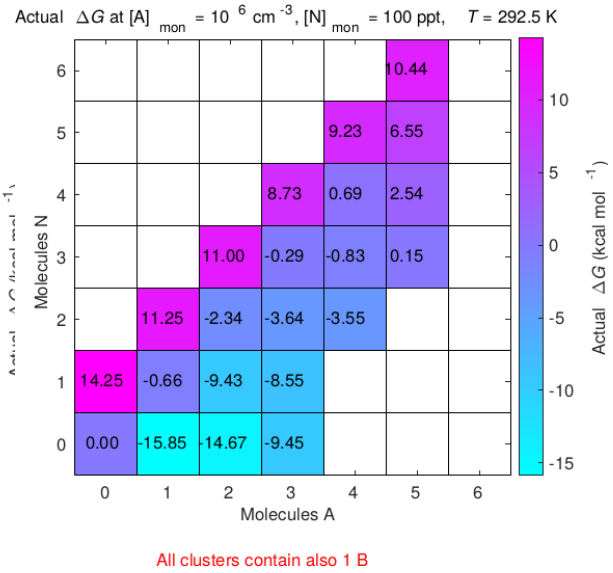
Figure A.1: A typical actual ΔG surface. $[\mathrm{A}]_{\mathrm{mon}}$ is the concentration of the sulfuric acid monomer, $[\mathrm{N}]_{\mathrm{mon}}$ the concentration of the ammonia monomer and B is bisulfate. Neutral mode depicts the saddle point shape. The critical cluster is the 4 sulfuric acid - 4 ammonia cluster. In charged modes most actual ΔGs are negative.



(a) neutral mode



(b) positive mode



(c) negative mode

Figure A.2: The actual ΔG surface for high temperature and low monomer concentrations. $[\mathrm{A}]_{\mathrm{mon}}$ is the concentration of the sulfuric acid monomer, $[\mathrm{N}]_{\mathrm{mon}}$ the concentration of the ammonia monomer and B is bisulfate. In neutral mode the actual ΔG continuously rises, there is no critical cluster. Even in charged mode actual ΔGs rise and even reach positive values.

Bibliography

[Baboul et al., 1999] Baboul, A. G., Curtiss, L. A., Redfern, P. C., and Raghavachari, K. (1999). Gaussian-3 theory using density functional geometries and zero-point energies. *J. Chem. Phys.*, 110(16):7650–7657.

[Bannwarth et al., 2019] Bannwarth, C., Ehlert, S., and Grimme, S. (2019). Gfn2-xtb-an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.*, 15(3):1652–1671.

[Bauschlicher and Partridge, 1995] Bauschlicher, C. W. and Partridge, H. (1995). A modification of the gaussian-2 approach using density functional theory. *J. Chem. Phys.*, 103(5):1788–1791.

[Becke, 1993] Becke, A. D. (1993). Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98(7):5648–5652.

[Bork et al., 2014] Bork, N., Elm, J., Olenius, T., and Vehkamäki, H. (2014). Methane sulfonic acid-enhanced formation of molecular clusters of sulfuric acid and dimethyl amine. *Atmos. Chem. Phys.*, 14(22):12023–12030.

[Born and Oppenheimer, 1927] Born, M. and Oppenheimer, R. (1927). Zur Quantentheorie der Molekeln. *Ann. Phys.*, 389(20):457–484.

[Brock et al., 1995] Brock, C. A., Hamill, P., Wilson, J. C., Jonsson, H. H., and Chan, K. R. (1995). Particle formation in the upper tropical troposphere: A source of nuclei for the stratospheric aerosol. *Science*, 270(5242):1650–1653.

[Capstick et al., 2015] Capstick, S., Whitmarsh, L., Poortinga, W., Pidgeon, N., and Upham, P. (2015). International trends in public perceptions of climate change over the past quarter century. *WIREs Clim. Change*, 6(1):35–61.

[Chai and Head-Gordon, 2008] Chai, J.-D. and Head-Gordon, M. (2008). Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.*, 10:6615–6620.

[Chaudhuri, 1994] Chaudhuri, B. (1994). How to choose a representative subset from a set of data in multi-dimensional space. *Pattern Recognit. Lett.*, 15(9):893 – 899.

[Clarke, 1992] Clarke, A. D. (1992). Atmospheric nuclei in the remote free-troposphere. *J. Atmos. Chem.*, 14(1):479–488.

[Constans et al., 2000] Constans, P., Ayala, P. Y., and Scuseria, G. E. (2000). Scaling reduction of the perturbative triples correction (t) to coupled cluster theory via laplace transform formalism. *J. Chem. Phys.*, 113(23):10451–10458.

[Dunne et al., 2016] Dunne, E. M., Gordon, H., Kürten, A., Almeida, J., Duplissy, J., Williamson, C., Ortega, I. K., Pringle, K. J., Adamov, A., Baltensperger, U., Barmet, P., Benduhn, F., Bianchi, F., Breitenlechner, M., Clarke, A., Curtius, J., Dommen, J., Donahue, N. M., Ehrhart, S., Flagan, R. C., Franchin, A., Guida, R., Hakala, J., Hansel, A., Heinritzi, M., Jokinen, T., Kangasluoma, J., Kirkby, J., Kulmala, M., Kupc, A., Lawler, M. J., Lehtipalo, K., Makhmutov, V., Mann, G., Mathot, S., Merikanto, J., Miettinen, P., Nenes, A., Onnela, A., Rap, A., Reddington, C. L. S., Riccobono, F., Richards, N. A. D., Rissanen, M. P., Rondo, L., Sarnela, N., Schobesberger, S., Sengupta, K., Simon, M., Sipilä, M., Smith, J. N., Stozkhov, Y., Tomé, A., Tröstl, J., Wagner, P. E., Wimmer, D., Winkler, P. M., Worsnop, D. R., and Carslaw, K. S. (2016). Global atmospheric particle formation from cern cloud measurements. *Science*, 354(6316):1119–1124.

[Elm, 2019] Elm, J. (2019). An atmospheric cluster database consisting of sulfuric acid, bases, organics, and water. *ACS Omega*, 4(6):10965–10974.

[Fock, 1930] Fock, V. (1930). Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Angew. Phys.*, 61(1):126–148.

[Frisch et al., 1990a] Frisch, M. J., Head-Gordon, M., and Pople, J. A. (1990a). A direct mp2 gradient method. *Chem. Phys. Lett.*, 166(3):275 – 280.

[Frisch et al., 1990b] Frisch, M. J., Head-Gordon, M., and Pople, J. A. (1990b). Semi-direct algorithms for the mp2 energy and gradient. *Chem. Phys. Lett.*, 166(3):281 – 289.

[Frisch et al., 2016] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V. G., Gao, J., Rega, N.,

Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J. J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Keith, T. A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M., Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B., and Fox, D. J. (2016). Gaussian 16 Revision C.01. Gaussian Inc. Wallingford CT.

[Funes-Ardoiz and Paton, 2016] Funes-Ardoiz, I. and Paton, R. S. (2016). Goodvibes: Goodvibes v1.0.1.

[Grimme, 2012] Grimme, S. (2012). Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chem. Eur. J.*, 18(32):9955–9964.

[Grimme et al., 2017] Grimme, S., Bannwarth, C., and Shushkov, P. (2017). A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (z = 1-86). *J. Chem. Theory Comput.*, 13(5):1989–2009.

[Hartree, 1928] Hartree, D. R. (1928). The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Math. Proc. Cambridge Philos. Soc.*, 24(1):89–110.

[Hättig and Weigend, 2000] Hättig, C. and Weigend, F. (2000). Cc2 excitation energy calculations on large molecules using the resolution of the identity approximation. *J. Chem. Phys.*, 113(13):5154–5161.

[Head-Gordon et al., 1988] Head-Gordon, M., Pople, J. A., and Frisch, M. J. (1988). Mp2 energy evaluation by direct methods. *Chem. Phys. Lett.*, 153(6):503 – 506.

[Henschel et al., 2016] Henschel, H., Kurtén, T., and Vehkamäki, H. (2016). Computational study on the effect of hydration on new particle formation in the sulfuric acid/ammonia and sulfuric acid/dimethylamine systems. *J. Phys. Chem. A*, 120(11):1886–1896.

[Hinds, 1999] Hinds, W. (1999). *Aerosol technology : properties, behavior, and measurement of airborne particles.* Wiley, New York.

[Hohenberg and Kohn, 1964] Hohenberg, P. and Kohn, W. (1964). Inhomogeneous Electron Gas. *Phys. Rev.*, 136:B864–B871.

[Jacobson et al., 2000] Jacobson, M. C., Hansson, H. C., Noone, K. J., and Charlson, R. J. (2000). Organic atmospheric aerosols: Review and state of the science. *Reviews of Geophysics*, 38(2):267–294.

[Jensen, 2007] Jensen, F. (2007). *Introduction to Computational Chemistry*. WILEY-VCH Verlag.

[Karaboga and Basturk, 2007] Karaboga, D. and Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. *J. Global Optim.*, 39(3):459–471.

[Kendall et al., 1992] Kendall, R. A., Dunning, T. H., and Harrison, R. J. (1992). Electron affinities of the first-row atoms revisited. systematic basis sets and wave functions. *J. Chem. Phys.*, 96(9):6796–6806.

[Kildgaard et al., 2018] Kildgaard, J. V., Mikkelsen, K. V., Bilde, M., and Elm, J. (2018). Hydration of atmospheric molecular clusters: A new method for systematic configurational sampling. *J. Phys. Chem. A*, 122(22):5026–5036.

[Klopper et al., 1997] Klopper, W., Noga, J., Koch, H., and Helgaker, T. (1997). Multiple basis sets in calculations of triples corrections in coupled-cluster theory. *Theor. Chem. Acc.*, 97(1):164–176.

[Kohn and Sham, 1965] Kohn, W. and Sham, L. J. (1965). Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, 140:A1133–A1138.

[Kubečka et al., 2019] Kubečka, J., Besel, V., Kurtén, T., Myllys, N., and Vehkamäki, H. (2019). Configurational sampling of noncovalent (atmospheric) molecular clusters: Sulfuric acid and guanidine. *J. Phys. Chem. A*, 123(28):6022–6033.

[Kulmala et al., 2004] Kulmala, M., Laakso, L., Lehtinen, K. E. J., Riipinen, I., Dal Maso, M., Anttila, T., Kerminen, V.-M., Hõrrak, U., Vana, M., and Tammet, H. (2004). Initial steps of aerosol growth. *Atmos. Chem. Phys.*, 4(11/12):2553–2560.

[Kupiainen-Määtä and Olenius, 2017] Kupiainen-Määtä, O. and Olenius, T. (2017). *ACDC: Atmospheric Cluster Dynamics Code Technical Manual*. University of Helsinki.

[Kürten et al., 2016] Kürten, A., Bianchi, F., Almeida, J., Kupiainen-Määttä, O., Dunne, E. M., Duplissy, J., Williamson, C., Barmet, P., Breitenlechner, M., Dommen, J., Donahue, N. M., Flagan, R. C., Franchin, A., Gordon, H., Hakala, J., Hansel, A., Heinritzi, M., Ickes, L., Jokinen, T., Kangasluoma, J., Kim, J., Kirkby, J., Kupc, A., Lehtipalo, K., Leiminger, M., Makhmutov, V., Onnela, A., Ortega,

I. K., Petäjä, T., Praplan, A. P., Riccobono, F., Rissanen, M. P., Rondo, L., Schnitzhofer, R., Schobesberger, S., Smith, J. N., Steiner, G., Stozhkov, Y., Tomé, A., Tröstl, J., Tsagkogeorgas, G., Wagner, P. E., Wimmer, D., Ye, P., Baltensperger, U., Carslaw, K., Kulmala, M., and Curtius, J. (2016). Experimental particle formation rates spanning tropospheric sulfuric acid and ammonia abundances, ion production rates, and temperatures. *J. Geophys. Res. D: Atmos.*, 121(20):12,377–12,400.

[Kurtén et al., 2007] Kurtén, T., Torpo, L., Ding, C.-G., Vehkamäki, H., Sundberg, M. R., Laasonen, K., and Kulmala, M. (2007). A density functional study on water-sulfuric acid-ammonia clusters and implications for atmospheric cluster formation. *J. Geophys. Res. D: Atmos.*, 112(D4).

[Leverentz et al., 2013] Leverentz, H. R., Siepmann, J. I., Truhlar, D. G., Loukonen, V., and Vehkamäki, H. (2013). Energetics of atmospherically implicated clusters made of sulfuric acid, ammonia, and dimethyl amine. *J. Phys. Chem. A*, 117(18):3819–3825.

[Li et al., 2015] Li, Y.-P., Gomes, J., Mallikarjun Sharada, S., Bell, A. T., and Head-Gordon, M. (2015). Improved force-field parameters for qm/mm simulations of the energies of adsorption for molecules in zeolites and a free rotor correction to the rigid rotor harmonic oscillator model for adsorption enthalpies. *J. Phys. Chem. C*, 119(4):1840–1850.

[Lohmann and Feichter, 2005] Lohmann, U. and Feichter, J. (2005). Global indirect aerosol effects: a review. *Atmos. Chem. Phys.*, 5(3):715–737.

[Loukonen et al., 2010] Loukonen, V., Kurtén, T., Ortega, I. K., Vehkamäki, H., Pádua, A. A. H., Sellegri, K., and Kulmala, M. (2010). Enhancing effect of dimethylamine in sulfuric acid nucleation in the presence of water - a computational study. *Atmos. Chem. Phys.*, 10(10):4961–4974.

[Mäkelä et al., 1997] Mäkelä, J. M., Aalto, P., Jokinen, V., Pohja, T., Nissinen, A., Palmroth, S., Markkanen, T., Seitsonen, K., Lihavainen, H., and Kulmala, M. (1997). Observations of ultrafine aerosol particle formation and growth in boreal forest. *Geophys. Res. Lett.*, 24(10):1219–1222.

[McGrath et al., 2012] McGrath, M. J., Olenius, T., Ortega, I. K., Loukonen, V., Paasonen, P., Kurtén, T., Kulmala, M., and Vehkamäki, H. (2012). Atmospheric cluster dynamics code: a flexible method for solution of the birth-death equations. *Atmos. Chem. Phys.*, 12(5):2345–2355.

[Merikanto et al., 2009] Merikanto, J., Spracklen, D. V., Mann, G. W., Pickering, S. J., and Carslaw, K. S. (2009). Impact of nucleation on global ccn. *Atmos. Chem. Phys.*, 9(21):8601–8616.

[Metzger et al., 2010] Metzger, A., Verheggen, B., Dommen, J., Duplissy, J., Prevot, A. S. H., Weingartner, E., Riipinen, I., Kulmala, M., Spracklen, D. V., Carslaw, K. S., and Baltensperger, U. (2010). Evidence for the role of organics in aerosol particle formation under atmospheric conditions. *PNAS*, 107(15):6646–6651.

[Møller and Plesset, 1934] Møller, C. and Plesset, M. S. (1934). Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618–622.

[Montgomery et al., 1999] Montgomery, J. A., Frisch, M. J., Ochterski, J. W., and Petersson, G. A. (1999). A complete basis set model chemistry. vi. use of density functional geometries and frequencies. *J. Chem. Phys.*, 110(6):2822–2827.

[Myhre et al., 2013] Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H. (2013). *Anthropogenic and natural radiative forcing*, pages 659–740. Cambridge University Press, Cambridge, UK.

[Myllys et al., 2016a] Myllys, N., Elm, J., Halonen, R., Kurtén, T., and Vehkamäki, H. (2016a). Coupled cluster evaluation of the stability of atmospheric acid-base clusters with up to 10 molecules. *J. Phys. Chem. A*, 120(4):621–630.

[Myllys et al., 2016b] Myllys, N., Elm, J., and Kurtén, T. (2016b). Density functional theory basis set convergence of sulfuric acid-containing molecular clusters. *Comput. Theor. Chem.*, 1098:1 – 12.

[Neese, 2018] Neese, F. (2018). Software update: the orca program system, version 4.0. *WIREs Comput. Mol. Sci.*, 8(1):e1327.

[Olenius et al., 2013a] Olenius, T., Kupiainen-Määttä, O., Ortega, I. K., Kurtén, T., and Vehkamäki, H. (2013a). Free energy barrier in the growth of sulfuric acid-ammonia and sulfuric acid-dimethylamine clusters. *J. Chem. Phys.*, 139(8):084312.

[Olenius et al., 2013b] Olenius, T., Schobesberger, S., Kupiainen-Määttä, O., Franchin, A., Junninen, H., Ortega, I. K., Kurtén, T., Loukonen, V., Worsnop, D. R., Kulmala, M., and Vehkamäki, H. (2013b). Comparing simulated and experimental molecular cluster distributions. *Faraday Discuss.*, 165:75–89.

[Ortega et al., 2012] Ortega, I. K., Kupiainen, O., Kurtén, T., Olenius, T., Wilkman, O., McGrath, M. J., Loukonen, V., and Vehkamäki, H. (2012). From quantum chemical formation free energies to evaporation rates. *Atmos. Chem. Phys.*, 12(1):225–235.

[Ortega et al., 2014] Ortega, I. K., Olenius, T., Kupiainen-Määttä, O., Loukonen, V., Kurtén, T., and Vehkamäki, H. (2014). Electrical charging changes the composition of sulfuric acid-ammonia/dimethylamine clusters. *Atmos. Chem. Phys.*, 14(15):7995–8007.

[Partanen et al., 2016] Partanen, L., Vehkamäki, H., Hansen, K., Elm, J., Henschel, H., Kurtén, T., Halonen, R., and Zapadinsky, E. (2016). Effect of conformers on free energies of atmospheric complexes. *J. Phys. Chem. A*, 120(43):8613–8624.

[Perdew and Wang, 1992] Perdew, J. P. and Wang, Y. (1992). Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45:13244–13249.

[Peverati and Truhlar, 2014] Peverati, R. and Truhlar, D. G. (2014). Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics. *Philos. Trans. R. Soc. London, Ser. A*, 372(2011):20120476.

[Pilati and Forni, 1998] Pilati, T. and Forni, A. (1998). Symmol.

[Pitoňák et al., 2006] Pitoňák, M., Holka, F., Neogrödy, P., and Urban, M. (2006). Optimized virtual orbitals for correlated calculations: Towards large scale ccsd(t) calculations of molecular dipole moments and polarizabilities. *J. Mol. Struct. THEOCHEM*, 768(1):79 – 89. Coupled-cluster Methods: Theory and Applications. A Collection of Invited Papers in Honor of Debashis Mukherjee on the Occasion of his 60th Birthday.

[Pople et al., 1993] Pople, J. A., Scott, A. P., Wong, M. W., and Radom, L. (1993). Scaling factors for obtaining fundamental vibrational frequencies and zero-point energies from hf/6-31g* and mp2/6-31g* harmonic frequencies. *Isr. J. Chem.*, 33(3):345–350.

[Riplinger and Neese, 2013] Riplinger, C. and Neese, F. (2013). An efficient and near linear scaling pair natural orbital based local coupled cluster method. *J. Chem. Phys.*, 138(3):034106.

[Riplinger et al., 2013] Riplinger, C., Sandhoefer, B., Hansen, A., and Neese, F. (2013). Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.*, 139(13):134101.

[Rose et al., 2015] Rose, C., Sellegri, K., Asmi, E., Hervo, M., Freney, E., Colomb, A., Junninen, H., Duplissy, J., Sipilä, M., Kontkanen, J., Lehtipalo, K., and Kulmala, M. (2015). Major contribution of neutral clusters to new particle formation at the interface between the boundary layer and the free troposphere. *Atmos. Chem. Phys.*, 15(6):3413–3428.

[Schobesberger et al., 2015] Schobesberger, S., Franchin, A., Bianchi, F., Rondo, L., Duplissy, J., Kürten, A., Ortega, I. K., Metzger, A., Schnitzhofer, R., Almeida, J., Amorim, A., Dommen, J., Dunne, E. M., Ehn, M., Gagné, S., Ickes, L., Junninen, H., Hansel, A., Kerminen, V.-M., Kirkby, J., Kupc, A., Laaksonen, A., Lehtipalo, K., Mathot, S., Onnela, A., Petäjä, T., Riccobono, F., Santos, F. D., Sipilä, M., Tomé, A., Tsagkogeorgas, G., Viisanen, Y., Wagner, P. E., Wimmer, D., Curtius, J., Donahue, N. M., Baltensperger, U., Kulmala, M., and Worsnop, D. R. (2015). On the composition of ammonia-sulfuric-acid ion clusters during aerosol particle formation. *Atmos. Chem. Phys.*, 15(1):55–78.

[Schobesberger et al., 2013] Schobesberger, S., Junninen, H., Bianchi, F., Lönn, G., Ehn, M., Lehtipalo, K., Dommen, J., Ehrhart, S., Ortega, I. K., Franchin, A., Nieminen, T., Riccobono, F., Hutterli, M., Duplissy, J., Almeida, J., Amorim, A., Breitenlechner, M., Downard, A. J., Dunne, E. M., Flagan, R. C., Kajos, M., Keskinen, H., Kirkby, J., Kupc, A., Kürten, A., Kurtén, T., Laaksonen, A., Mathot, S., Onnela, A., Praplan, A. P., Rondo, L., Santos, F. D., Schallhart, S., Schnitzhofer, R., Sipilä, M., Tomé, A., Tsagkogeorgas, G., Vehkamäki, H., Wimmer, D., Baltensperger, U., Carslaw, K. S., Curtius, J., Hansel, A., Petäjä, T., Kulmala, M., Donahue, N. M., and Worsnop, D. R. (2013). Molecular understanding of atmospheric particle formation from sulfuric acid and large oxidized organic molecules. *PNAS*, 110(43):17223–17228.

[Shampine and Reichelt, 1997] Shampine, L. F. and Reichelt, M. W. (1997). The matlab ode suite. *SIAM J. Sci. Comput.*, 18(1):1–22.

[Stephens et al., 1994] Stephens, P. J., Devlin, F. J., Chabalowski, C. F., and Frisch, M. J. (1994). Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.*, 98(45):11623–11627.

[Sun and Ariya, 2006] Sun, J. and Ariya, P. A. (2006). Atmospheric organic and bioaerosols as cloud condensation nuclei (ccn): A review. *Atmos. Environ.*, 40(5):795 – 820.

[Sure et al., 2014] Sure, R., Antony, J., and Grimme, S. (2014). Blind prediction of binding affinities for charged supramolecular host–guest systems: Achievements and shortcomings of dft-d3. *J. Phys. Chem. B*, 118(12):3431–3440.

[Vanommeslaeghe et al., 2010] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and Mackerell Jr., A. D. (2010). Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *J. Comput. Chem.*, 31(4):671–690.

[Virtanen et al., 2010] Virtanen, A., Joutsensaari, J., Koop, T., Kannosto, J., Yli-Pirilä, P., Leskinen, J., Mäkelä, J. M., Holopainen, J. K., Pöschl, U., Kulmala, M., Worsnop, D. R., and Laaksonen, A. (2010). An amorphous solid state of biogenic secondary organic aerosol particles. *Nature*, 467(7317):824–827.

[Vosko et al., 1980] Vosko, S. H., Wilk, L., and Nusair, M. (1980). Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.*, 58(8):1200–1211.

[Wang and Penner, 2009] Wang, M. and Penner, J. E. (2009). Aerosol indirect forcing in a global model with particle nucleation. *Atmos. Chem. Phys.*, 9(1):239–260.

[Yu and Luo, 2009] Yu, F. and Luo, G. (2009). Simulation of particle size distribution with a global aerosol model: contribution of nucleation to aerosol and ccn number concentrations. *Atmos. Chem. Phys.*, 9(20):7691–7710.

[Yu et al., 2012] Yu, W., He, X., Vanommeslaeghe, K., and MacKerell Jr., A. D. (2012). Extension of the charmm general force field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput. Chem.*, 33(31):2451–2468.

[Zhang et al., 2018] Zhang, H., Li, H., Liu, L., Zhang, Y., Zhang, X., and Li, Z. (2018). The potential role of malonic acid in the atmospheric sulfuric acid - ammonia clusters formation. *Chemosphere*, 203:26 – 33.

[Zhang and Dolg, 2015] Zhang, J. and Dolg, M. (2015). Abcluster: the artificial bee colony algorithm for cluster global optimization. *Phys. Chem. Chem. Phys.*, 17:24173–24181.

[Zhang and Dolg, 2016] Zhang, J. and Dolg, M. (2016). Global optimization of clusters of rigid molecules using the artificial bee colony algorithm. *Phys. Chem. Chem. Phys.*, 18:3003–3010.